

Катя Рашева-Йорданова
Стефка Толева-Стоименова

ИЗГРАЖДАНЕ НА DATA SCIENCE КОМПЕТЕНТНОСТ
ЗА ПРЕОДОЛЯВАНЕ НА ДИГИТАЛНОТО НЕРАВЕНСТВО

Katia Rasheva-Yordanova
Stefka Toleva-Stoimenova

BUILDING DATA SCIENCE COMPETENCE
FOR OVERCOMING THE DIGITAL DIVIDE

Катя Рашева-Йорданова
Стефка Толева-Стоименова

**ИЗГРАЖДАНЕ
НА DATA SCIENCE КОМПЕТЕНТНОСТ
ЗА ПРЕОДОЛЯВАНЕ
НА ДИГИТАЛНОТО НЕРАВЕНСТВО**

Академично издателство „За буквите – О писменехъ“
София, 2019



Монографията е изготвена с финансовата помощ
на Фонд „Научни изследвания“.
Проект „Изграждане на Data Science компетентност за преодоляване
на дигиталното неравенство“ Договор: ДМ 12/04 от 20.12.2017 г.

© Катя Христова Рашева-Йорданова, автор, 2019
© Стефка Янкова Толева-Стоименова, автор, 2019
© Доц. д-р Волин Карагъзов, научен рецензент, 2019
© Доц. д-р Евгения Ковачева, научен рецензент, 2019
© Ралица Йотова, графичен дизайн на корицата, 2019
© Академично издателство „За буквите – О писменехъ“
ISBN 978-619-185-392-2

София, 2019

© Katia Rasheva-Yordanova, author, 2019
© Stefka Toleva-Stoimenova, author, 2019
© Assoc. Prof. Volin Karagyozov, PhD, scientific reviewer, 2019
© Assoc. Prof. Eugenia Kovatcheva, PhD, scientific reviewer, 2019
© Ralitsa Iotova, graphic design of the cover, 2019
© Academic Publishing House “Za Bukvite – O Pismeneh”
ISBN 978-619-185-392-2

Sofia, 2019

Авторките изказват дълбоката си благодарност за позитивната и конструктивна критика, насоките, полезните съвети и постоянната подкрепа при осъществяването на проекта и работата по настоящата монография на проф. д.т.н. Димитър Христов.

СЪДЪРЖАНИЕ

ВЪВЕДЕНИЕ	19
ГЛАВА I.	
ОСНОВОПОЛАГАЩИ ПОНЯТИЯ.....	23
1.1. ГОЛЕМИ ДАННИ.....	23
1.1.1. Характеристики на големите данни	23
1.1.2. Дефиниции.....	26
1.1.3. Проблеми и предизвикателства	28
1.1.4. Съхранение и обработка на големи данни.....	30
1.1.5. Обработка и анализ на големи данни	33
1.2. НАУКА ЗА ДАННИТЕ.....	38
1.2.1. Основни дефиниции.....	38
1.2.2. Мулти-, интер- и трансдисциплинарност на науката за данните	41
1.2.3. Процес на информационно осигуряване.....	49
1.1.1. Основни приложения.....	52
1.3. ДИГИТАЛНО НЕРАВЕНСТВО В КОНТЕКСТА НА ГОЛЕМИТЕ ДАННИ	53
1.3.1. Грамотност за работа с големи данни	53
1.3.2. Дигитално неравенство	59
1.3.3. Измерения и измерители.....	61
1.3.4. Неравенство по отношение на науката за данните	65
1.3.5. Подходи за преодоляване на неравенството по отношение на големите данни.....	68
РЕЗУЛТАТИ И ИЗВОДИ ПО ПЪРВА ГЛАВА	71
ГЛАВА II.	
ПРЕГЛЕД НА ЗАДЪЛЖИТЕЛНИТЕ КОМПЕТЕНЦИИ В ЕРАТА НА ГОЛЕМИТЕ ДАННИ.....	73
2.1. КОМПЕТЕНТНОСТЕН ПОДХОД	73
2.2. ПРОФЕСИОНАЛЕН ПРОФИЛ НА СПЕЦИАЛИСТА ПО ДАННИ.....	78
2.2.1. Специалист по данни	78
2.2.2. Рамка на ключовите компетентности.....	86

2.2.3. Аналитични умения.....	90
2.2.4. Етични умения.....	102
2.3. ПОТРЕБНОСТ ОТ ЕКСПЕРТИЗА В ОБЛАСТТА НА НАУКАТА ЗА ДАННИ	109
2.3.1. Търсене на специалисти по данни в световен мащаб	109
2.3.1. Търсене на специалисти по данни в България.....	111
РЕЗУЛТАТИ И ИЗВОДИ ПО ГЛАВА II.....	118
ГЛАВА III.	
ОБУЧЕНИЕ В ОБЛАСТТА НА НАУКАТА ЗА ДАННИТЕ	119
3.1. ПРОЦЕС НА ОБУЧЕНИЕ В ОБРАЗОВАТЕЛНИТЕ ИНСТИТУЦИИ.....	119
3.1.1. Таксономии за целите на образованието	119
3.1.2. Предизвикателства пред обучаващите институции.....	123
3.1.3. Обучение, базирано на компетенции.....	127
3.1.4. Разработване на компетенциите	129
3.2. ПРЕГЛЕД НА ПРОГРАМИТЕ В ОБЛАСТТА НА НАУКАТА ЗА ДАННИТЕ.....	136
3.2.1. Състояние и тенденции за развитие в световен мащаб.....	136
3.2.2. Състояние на обучението по науката за данните в България.....	145
3.3. ПРОУЧВАНЕ НА ВХОДЯЩИТЕ УМЕНИЯ НА ПОТЕНЦИАЛНИТЕ КАНДИДАТИ ЗА МАГИСТЪРСКА ПРОГРАМА ПО НАУКА ЗА ДАННИТЕ	149
3.3.1. Аналитичното мислене като основна компетентност.....	149
3.3.2. Разработване на въпросник	151
3.3.3. Анализ на събраните от проучването данни	155
РЕЗУЛТАТИ И ИЗВОДИ ПО ГЛАВА III	161
ГЛАВА IV.	
ИЗГРАЖДАНЕ НА МОДЕЛ НА УЧЕБЕН ПЛАН, БАЗИРАН НА КОМПЕТЕНЦИИ	163
4.1. ФОРМИРАНЕ НА КОМПЕТЕНЦИИ В ОБЛАСТТА НА НАУКАТА ЗА ДАННИТЕ.....	164
4.1.1. Методика за разработване на учебна програма, основана на компетенции	164
4.1.2. Общ модел на система за формиране на професионални компетенции в областта на науката за данните	169

4.1.3. Потенциални кандидати за магистърската програма	174
4.2. ПРОЕКТИРАНЕ НА МАГИСТЪРСКА ПРОГРАМА „АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“	180
4.2.1. Проектиране на учебен план – предпоставки и нормативна уредба	180
4.2.2. Учебна документация	184
4.2.3. Управление на риска при обучението по анализ на големи данни	190
4.3. КУРС ПО „ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ“	194
4.3.1. Проектиране на учебна програма на дисциплината	194
4.3.2. Методология за обучение по „Въведение в науката за данните“	197
РЕЗУЛТАТИ И ИЗВОДИ ПО ГЛАВА IV	208
ЗАКЛЮЧЕНИЕ	209
ОСНОВНИ ПРИНОСИ В МОНОГРАФИЯТА	211
Приложение 1. АНКЕТНА КАРТА	213
Приложение 2. АЛГОРИТЪМ НА ОБРАБОТКА НА ДАННИТЕ	217
Приложение 3. КАТАЛОГ ОТ КОМПЕТЕНЦИИ КЪМ ПРОГРАМА „АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“	220
Приложение 4. УЧЕБЕН ПЛАН НА МАГИСТЪРСКА ПРОГРАМА „АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“	223
Приложение 5. КВАЛИФИКАЦИОННА ХАРАКТЕРИСТИКА НА СПЕЦИАЛНОСТ „АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“	230
Приложение 6. УЧЕБНА ПРОГРАМА ПО ДИСЦИПЛИНАТА „ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИ (DATA SCIENCE)“	233
Приложение 7. УЧЕБНА МЕТОДОЛОГИЯ	236
СПИСЪК НА ИЗПОЛЗВАНИТЕ СЪКРАЩЕНИЯ	246
ТЕРМИНОЛОГИЧЕН РЕЧНИК	248
СПИСЪК НА ФИГУРИТЕ	250
СПИСЪК НА ТАБЛИЦИТЕ	252

ИЗПОЛЗВАНА ЛИТЕРАТУРА	255
ПУБЛИКАЦИИ, РЕАЛИЗИРАНИ ПО ПРОЕКТ ДМ 12/04 ОТ 20.12.2017 г., ИЗПОЛЗВАНИ В МОНОГРАФИЯТА.....	275
Рецензия. Доц. д-р Волин Карагъзов.....	277
Рецензия. Доц. д-р Евгения Ковачева.....	279

CONTENTS

INTRODUCTION.....	15
CHAPTER I. BASIC NOTIONS.....	23
1.1. BIG DATA	23
1.1.1. Big Data Characteristics.....	23
1.1.2. Definitions	26
1.1.3. Issues and Challenges.....	28
1.1.4. Big Data Storage and Processing.....	30
1.1.5. Processing and Analysis.....	33
1.2. DATA SCIENCE.....	38
1.2.1. Basic Definitions.....	38
1.2.2. Multi-, Inter- and Trans-Disciplinarity of the Data Science	41
1.2.3. The Process of Information Provision.....	49
1.2.4. Basic Applications	52
1.3. THE DIGITAL DIVIDE IN THE CONTEXT OF BIG DATA	53
1.3.1. Literacy for Working with Big Data.....	53
1.3.2. The Digital Divide	59
1.3.3. Measures and Measurements.....	61
1.3.4. Data Science Divide	65
1.3.5. Approaches to Overcoming the Big Data Divide	68
CHAPTER I RESULTS AND CONCLUSIONS.....	71
CHAPTER II. OVERVIEW OF THE COMPULSORY COMPETENCES IN THE ERA OF BIG DATA	73
2.1. THE COMPETENCE APPROACH.....	73
2.2. PROFESSIONAL PROFILE OF THE DATA SPECIALIST... ..	78
2.2.1. Data Specialist.....	78
2.2.2. Framework of Key Competences.....	86
2.2.3. Analytical Skills	90
2.2.4. Ethical Skills	102

2.3. THE NEED FOR EXPERTISE IN THE FIELD OF DATA SCIENCE	109
2.3.1. The Demand for Data Specialists Worldwide	109
2.3.2. The Demand for Data Specialists in Bulgaria.....	111
CHAPTER II RESULTS AND CONCLUSIONS	118
3. CHAPTER III. TRAINING IN THE FIELD OF DATA SCIENCE.....	119
3.1. TRAINING PROCESSES IN EDUCATIONAL INSTITUTIONS.....	119
3.1.1. Taxonomies on the Aims of Education	119
3.1.2. Challenges Facing Training Institutions	123
3.1.3. Competence-Based Training.....	127
3.2. OVERVIEW OF CURRICULA IN THE FIELD OF DATA SCIENCE	136
3.2.1. Current Situation and Development Trends Worldwide	136
3.2.2. Current Situation of the Training in Data Science in Bulgaria.....	145
3.3. A SURVEY ON THE INPUT SKILLS OF POTENTIAL CANDIDATES FOR A MASTER'S PROGRAM IN DATA SCIENCE	149
3.3.1. Analytical Thinking as a Basic Competence.....	149
3.3.2. Developing a Questionnaire.....	151
3.3.3. Analysis of the Survey Data	155
CHAPTER III RESULTS AND CONCLUSIONS.....	161
CHAPTER IV. CREATING A MODEL FOR A COMPETENCE-BASED CURRICULUM.....	163
4.1. COMPETENCE FORMING IN THE FIELD OF DATA SCIENCE	164
4.1.1. Methodology for Developing a Competence-Based Curriculum	164
4.1.2. A common Model for a System for Forming Professional Competences in the Field of Data Science	169
4.1.3. Potential Master's Program Students	174
4.2. DESIGNING A MASTER'S PROGRAM "DATA ANALYSIS AND MANAGEMENT".....	180
4.2.1. Designing a Curriculum – Prerequisites and Normative Regulation	180
4.2.2. Learning Documentation.....	184

4.2.3. Risk Management as an Essential Component of Training in Big Data Analysis	190
4.3. “INTRODUCTION TO DATA SCIENCE” COURSE	194
4.3.1. Designing a Course Syllabus	194
4.3.2. Methodology of Training in “Introduction to Data Science” ..	197
CHAPTER IV RESULTS AND CONCLUSIONS	208
CONCLUSION.....	209
MAJOR CONTRIBUTIONS OF THE MONOGRAPH.....	211
LIST OF APPENDICES	213
Appendix 1. SURVEY QUESTIONNAIRE.....	213
Appendix 2. ALGORITHM FOR DATA PROCESSING.....	217
Appendix 3. CATALOG OF COMPETENCES IN THE MASTER’S PROGRAM “DATA ANALYSIS AND MANAGEMENT”	220
Appendix 4. CURRICULUM OF THE MASTER’S PROGRAM “DATA ANALYSIS AND MANAGEMENT”	223
Appendix 5. QUALIFICATIN CHARACTERISTICS OF THE SPECIALTY “DATA ANALYSIS AND MANAGEMENT”	230
Appendix 6. A SYLLABUS OF DATA SCIENCE COURSE	233
Appendix 7. TEACHING METHODOLOGY	236
LIST OF ABBREVIATIONS.....	246
GLOSSARY OF TERMS.....	248
LIST OF FIGURES.....	250
LIST OF TABLES	252
REFERENCES	255
LIST OF PUBLICATIONS USED IN THE MONOGRAPH AND FUNDED BY PROJECT ΔM 12/04 OF 20.12.2017	275
Review. Assoc. Prof. Volin Karagiozov, PhD	277
Review. Assoc. Prof. Eugenia Kovatcheva, PhD	279

INTRODUCTION

With the advent of the Big Data phenomenon, there has been a significant increase in the demand for professionals with the competencies and skills to use big data efficiently and effectively. This monograph examines a new phase of the digital divide between the vast majority of society and the elite, who are able to benefit from informed, data-driven decision making.

This paper presents the main results of the research project DM 12/4 of December 20, 2017, entitled “Formation of Data Science Competence for Overcoming the Digital Divide”, funded by the Bulgarian Science Fund.

The primary **objective** of the monograph is to present the creation of a Data Science training plan as an approach to overcoming the digital divide in the Big Data Divide.

The tasks that follow as a result of the objective are:

1. To identify the framework of competences in the field of Data Science on the basis of a literature review and analysis of world practices;
2. To study the skills that the modern data specialist must possess, identified as key to successful implementation in the field and sought by organizations;
3. To analyze the demand and supply of specialists in the field of Data Science in Bulgaria.
4. To develop a curriculum for training in a Master's degree program in Data Science, tailored simultaneously to the needs of organizations and to the prior preparation of potential applicants for training.

The methods in the current work include:

- Literature analysis - publications on the problem in Bulgaria and abroad, academic, applied, labor market research statistics and forecasts for its development..
- Analysis of job advertisements on major Bulgarian sites.
- Direct interviews with employers seeking suitable professionals and with industry representatives providing such services.
- Modeling related to the application of competence-based teaching in pedagogical practice.
- Interviewing students.

- Mathematical and statistical methods for analysis of quantitative and qualitative data and verification of dependencies.

The scope of the research is determined by the set tasks, the objective being to propose an educational model to meet the needs of the Bulgarian labor market and the demand for data analysts from the Bulgarian industry.

Significance of the research

The significance of the problems addressed in this monograph is expressed in its practical focus. The study examined and identified the skills that a data analyst (Data Scientist) should possess. The study covers both literature research on the subject and business needs in Bulgaria. The study focuses on 4 categories of skills: hard skills, soft skills, analytical skills and ethical skills.

The curriculum of the Master's program in Data Science has been developed in response to the demand for Big Data professionals on the Bulgarian market. The training program includes disciplines that build the necessary skills and competencies of future data professionals, taking into account market needs, as well as generally accepted higher education practices. Building skills also implies the fitting of certain pedagogical practices.

Developing competences related to data science activities is still a major challenge. This is mainly due to the lack of a single framework to guide the formation of the professional profile of data experts, as well as the lack of historical experience in training professionals with the necessary competencies to be successful in the long term in a fast changing environment.

Research periods

The overall study has been conducted in several stages. The theoretical and conceptual stage is connected with the study of literary sources on the problem in order to define the main terms and issue areas. In the course of the project, the concept of the research was developed, theoretical information and empirical data were accumulated for operationalization of the competence model of students in the specialty “Data Analysis and Management” at the Faculty of Information Sciences at ULSIT (the University of Library Studies and Information Technologies). This is reflected in the implementation of the first, second and third research tasks.

The experimental stage comprises modeling the system for the formation of professional competences in the field of Data Science. The stage also includes

the creation of a curriculum for the Master's program in Data Analysis and Management, focused on building professional competencies for working with Big Data. The 4th research task is in focus here.

The final stage is related to the summary of the results of the study and the emphasis on parts of it. Summary and conclusions have been formulated, areas for future research have been outlined.

As an **experimental base** for the research has been used the Faculty of Information Sciences at ULSIT.

Structure of the work

Structurally, the monograph consists of an introduction, four chapters, a conclusion, a list of used literature, a glossary, a list of figures, and a list of tables.

The first chapter is devoted to clarifying the framework of research by defining the basic terms and building the conceptual apparatus. The chapter discusses the context and major challenges facing the industry in the current stage of information and communication technologies development, related to the accumulation of large amounts of data in computer systems and the need to benefit from them. These new conditions have led to a new stage of globalization and a new type of competitive challenge. The social aspects of this new phase of “digital divide” – the inequality of reaping the benefits of big data – have also been examined.

Chapter Two provides an overview of the literature and other information sources regarding the requirements for a specialist data analyst. The analysis has allowed us to identify the main categories of competences and skills that from today's point of view a data analyst must possess and which are recognized by employers.

Chapter three looks at the different educational approaches and practices in training data analysts and addresses key aspects related to training students in data science programs. The chapter presents the results of an empirical study concerning the analytical skills possessed by potential applicants for the Master's program in Data Analysis and Management.

Chapter four summarizes the results of the study and provides a model for a curriculum in data science. The proposed Data Analysis and Management curriculum is based on the competence model and the formation of the necessary competences for future data professionals.

ВЪВЕДЕНИЕ

С появата на феномена „големи данни“ (Big Data) значително нарасна търсенето на специалисти, притежаващи компетентности и умения за ефективното и полезното им използване. Настоящата монография разглежда нова фаза на дигитално разделение между огромното мнозинство от обществото и „елита“, способен да се възползва от вземането на информирани решения, основани на големи данни.

В настоящия труд са представени основните резултати, реализирани по линия на финансирания от Фонд „Научни изследвания“ научноизследователски проект ДМ 12/4 от 20.12.2017 г. на тема „Формиране на Data Science компетентност за преодоляване на дигиталното неравенство“.

Основната **цел** на монографията е да представи създаването на план за обучение по наука за данните (Data Science) като подход за преодоляване на дигиталното разделение (Big Data Divide) в епохата на големите данни.

Задачите, които произтичат от поставената цел, са следните:

1. Да се идентифицира рамката на компетентностите в областта на Data Science въз основа на литературен преглед и анализ на световни практики;
2. Да се проучат уменията, които трябва да притежава съвременният специалист по данни, които се идентифицират като ключови за успешна реализация в областта и са търсени от организациите;
3. Да се направи анализ на търсенето и предлагането на специалисти в областта на науката за данните в България;
4. Да се разработи учебен план за обучение в магистърска програма по наука за данните, който да е съобразен едновременно с нуждите на организациите и с предварителната подготовка на потенциалните кандидати за обучението.

Методите, приложени в настоящия труд, включват:

- анализ на литература по темата – публикации по проблема у нас и в чужбина, академични и приложни статистики от изследвания на пазара на труда и прогнози за развитието му;
- анализ на обявите за работа в основни български сайтове;
- директни интервюта с работодатели, търсещи подходящи

специалисти;

- моделиране, свързано с прилагането в педагогическата практика на обучение, базирано на компетенции;
- анкетиране на студенти;
- математико-статистически методи за анализ на количествени и качествени данни и проверка на зависимости.

Обхватът на изследването се определя от поставените задачи, като целта е да се предложи образователен модел, който да отговори на потребностите на българския пазар на труда и търсенето на анализатори на данни от българската индустрия.

Значимостта на засегнатите в настоящата монография проблеми се изразява в практическата ѝ насоченост. В изследването са проучени и идентифицирани уменията, които трябва да притежава специалистът, анализиращ (големи) данни (Data Scientist). Проучването обхваща както изследване на литература по въпроса, така и проучване на потребностите на бизнеса в България. Ориентирано е към 4 категории умения: „твърди“ умения, „меки“ умения, аналитични умения и етични умения.

Учебният план на магистърската програма по наука за данните е разработен в отговор на потребността от професионалисти в тази област на българския пазар. Обучителната програма включва изучаването на дисциплини, които изграждат необходимите умения и компетентности на бъдещите специалисти по данни, като се отчита пазарната готовност, както и общоприети практики във висшето образование. Изграждането на съответните умения предполага и прилагането на определени педагогически практики.

Развитието на компетентностите, свързани с дейностите в областта на науката за данните, все още представлява голямо предизвикателство. Това се дължи основно на липсата на единна рамка, която да служи като насока за формиране на професионалния профил на специалистите по данни, а също така и на липсата на исторически опит в обучението на професионалисти с необходимите компетентности, които да са успешни в дългосрочен план в условията на бързо променяща се среда.

Цялостното изследване се провежда в няколко **етапа**. *Теоретико-концептуалният* етап се свързва с проучване на литературни източници по проблема с цел определяне на основните термини и проблемни области. В хода му се изгражда концепцията за изследването, натрупват се теоретична информация и емпирични данни за операционализация на компетентностния модел на студентите по специалността

„Анализ и управление на данни“ във Факултет „Информационни науки“ към Университета по библиотекознание и информационни технологии (УниБИТ). Реализират се първите три изследователски задачи.

Експерименталният етап обхваща моделирането на системата за формиране на професионални компетенции в областта на науката за данните. Етапът включва и създаването на учебен план на магистърска програма „Анализ и управление на данни“, която е ориентирана към изграждането на професионални компетенции за работа с големи данни. Изпълнява се четвъртата от изследователските задачи.

Заклучителният етап се свързва с обобщаване на получените от изследването резултати и извеждане на акценти от него. Формулират се обобщения и изводи, очертават се направления за бъдещи научни изследвания.

Като **експериментална база** на изследването се използва Факултет „Информационни науки“ на УниБИТ.

В **структурно отношение** монографията се състои от въведение, четири глави, заключение, списък на използваната литература, терминологичен речник, списък на приложените фигури и таблици.

В глава I се изяснява рамката на изследването, дефинират се основните термини и се изгражда понятийният апарат. Коментират се контекстът и основните предизвикателства пред индустрията в условията на съвременния етап на развитие на информационните и комуникационните технологии, който е свързан с натрупването на големи масиви от данни в компютърните системи и необходимостта от извличане на полза от тях. Тези нови условия водят до нов етап на глобализация и нов тип предизвикателства за осигуряване на конкурентност. В главата са разгледани и социалните аспекти на този нов етап на „дигитално неравенство“ – неравенство по отношение на извличането на полза от големи данни.

В глава II е направен обзор на литературата и другите информационни източници по отношение на изискванията към специалиста – анализатор на данни. Проучването им позволява да се определят основните категории компетентности и умения, които, от днешна гледна точка, трябва да притежава анализаторът на данни и които работодателите изискват.

В глава III са разгледани различните образователни подходи и практики в обучението на анализаторите на данни и са засегнати основни аспекти, свързани с обучението на студенти по програми в областта на науката за данните. Представени са резултатите от извърше-

но емпирично изследване, засягащо аналитичните умения, които трябва да притежават потенциалните кандидати за магистърска програма „Анализ и управление на данни“.

В глава IV е направено обобщение на резултатите от изследването и е предложен модел на учебна програма за обучение по наука за данните. Разгледаната учебна програма по „Анализ и управление на данни“ се базира на компетентностния модел и формирането на необходимите компетенции на бъдещите специалисти по данни.

ГЛАВА I.

ОСНОВОПОЛАГАЩИ ПОНЯТИЯ

През последните години се наблюдава експоненциално нарастване на обема на генерираните от хората и машините данни. Появи се парадигмата на „големите данни“ (Big Data). Те представляват огромни масиви от данни с различен характер и висок темп на изменение, които изискват използването на нови начини и технологии за събиране, обработка и извличане на полезни знания. Това наложи ново тълкуване на проблемите, свързани с данните. Утвърди се нова мулти- и интердисциплинарна област, популярна като наука за данните (Data Science), в която се включват статистика, информатика, компютърни технологии, комуникация, управление, социология и др. Тя се занимава с анализ на структурирани или неструктурирани данни и извличане на скрити смислени модели, вътрешни връзки и зависимости, които не може да бъдат непосредствено установени от натрупаните данни. Днес науката за данните е твърде обхватна сфера, която намира много и разнообразни приложения в различни области на стопанския и обществения живот. Анализът на големите данни позволява на анализаторите, изследователите и бизнес потребителите да вземат по-ефективни решения чрез използването на данни, които преди това са били недостъпни, забудени или неизползваеми. От друга страна, компетентността за работа с данни и владенето на инструментариума на науката за данните създават нова разделителна линия между „елита“ и останалите членове на обществото, която се явява естествена нова фаза в наблюдаваното от години дигитално неравенство.

1.1. ГОЛЕМИ ДАННИ

1.1.1. Характеристики на големите данни

За първи път през 2001 г. Laneу [139] споменава трите V-та (Volume, Variety, Velocity), които са в основата на определянето на „големите данни“ (Big Data). Обемът (Volume) се отнася до огромния физически обем на масивите от данни, скоростта на нарастване (Velocity) – до актуалността, скоростта на обновяване, обработка и получаване на резултат в реално време, а многообразието (Variety) е свързано с различните източници и форми на представяне на данните.

Други автори се придържат към дефиницията на петте V-та, която добавя стойност (*Value*) и достоверност (*Veracity*) към „3V“ модела. [78, 103, 128, 153]. Напоследък някои изследователи дори използват повече V-та, за да опишат структурните характеристики на големите данни [79, 174]. На фиг. 1.1 е представен обобщен „5V“ модел на големите данни.



Фиг. 1.1. „5V“ модел на големите данни (адаптирана от Younas, 2018 [239])

Ще разгледаме подробно всяка една от характеристиките на големите данни, включени в дадения модел [240]:

- **Обем:** Тази характеристика се отнася до огромното количество данни, които се генерират, събират и обработват (в размер на петабайтове, екзбайтове и зетабайтове). Количеството налични и достъпни данни далеч надхвърля човешките способности за директна работа с тях. Например социалните мрежи Twitter и Facebook обработват милиони публикации и изображения, а Google получава над 63 000 търсения в секунда всеки ден. Освен това милиони записи на данни, свързани с транспорта, времето, системите за опазване на околната среда и др., се събират от сензорни технологии.

- **Скорост:** Тази характеристика се свързва със скоростта на генериране, обработка и трансфер на данни между различните системи и устройства, а също и с динамиката на актуализация на данните. Примерите включват скорост на публикациите в социалните медии, онлайн трансакции и проверка на измами, ежеминутни промени в данните за превоз на автобуси, влакове, самолети и др.

- **Разнообразие:** Тази характеристика се отнася до различните типове данни (структурирани, полуструктурирани и неструктурирани), които може да бъдат използвани в единство за извличане на полезното знание от тях. Разнообразието се свързва с променливостта на обстоятелствата в генерираните данни, но също и с различните начини за записване на данните.

- **Истинност:** Отнася се до качеството на данните, като коректност, последователност, доверие, сигурност и надеждност. Неспособността да се проверят данните в настоящата, много замърсена, информационна среда, не допринася за повишаване на достоверността на получената информация. Други проблеми и предизвикателства са свързани с неразбирането или неправилното интерпретиране на информацията, както и с липсата на компетенции по отношение на достоверността на данните.

- **Стойност:** Отнася се до различните видове ползи, които може да бъдат получени от обработката и анализирането на големи данни. Примерите включват: парична стойност; социална стойност; стойност на изследването/образованието и др.

Figarо дефинира основните предизвикателства и предимства на прилагането на големи данни чрез 10 *V*-та. Освен гореспоменатите 5 основни *V*-та той включва следните характеристики [101]:

- **Променливост** (*Variability*) – свързана е с аномалии, отклонения и други несъответствия в данните, а също и с непоследователната скорост, с която големите данни се зареждат в базата данни.

- **Валидност** (*Validity*) – подобно на истинността, валидността се отнася до точността и коректността на данните.

- **Уязвимост** (*Vulnerability*) – отнася се до рисковете, свързани със сигурността, тъй като много често големите данни са свързани с лична и чувствителна информация.

- **Колебливост** (*Volatility*) – свързва се с динамичното поведение и развитие на големите данни във времето и пространството.

- **Визуализация** (*Visualization*) – поради спецификата на големите данни възниква необходимост от нови, различни начини за ин-

тегриране, обработка и онагледяване на резултатите.

Напоследък се използват и много други характеристики на големите данни. През 2017 г. Tom Shafer съставя списък с 42 *V*-та, в който допълва важни характеристики, като: **виралност** (*Virality*), **местоположение** (*Venue*), **речник** (*Vocabulary*), **неопределеност** (*Vagueness*) и др. [204]. Необходимо е да се отбележи, че редът на подредба на характеристиките на големите данни не е редът им на значимост. В различните приложения приоритетите са различни. Дефиницията на Christozov, Toleva [55] търси по-абстрактно характеризирани на големите данни, като избягва изброяването на отделните характеристики, формиращи сложността на този феномен, а също и зависимостта от моментното състояние на технологиите. Те дефинират големите данни като горната граница на сложност на данните, които анализаторите могат ефективно да анализират с помощта на достъпните към момента информационни технологии. Така задачата, решена от Hollerith през 90-те години на 19. век, за обработка на данните за около 60 милиона американци при второто национално преброяване на населението е представлявала далеч по-голямо предизвикателство за времето си от обработването на 100 и повече терабайта информация днес.

1.1.2. Дефиниции

Днес терминът „големи данни“ се използва за огромни масиви от данни, които е трудно, дори невъзможно да бъдат обработени с познатите методи и технологии за работа с информационни източници. Съществуват множество различни дефиниции на термина, които често взаимно се припокриват или допълват. В Таблица 1.1 са посочени дефиниции, добили популярност в научната литература, които дават различни гледни точки за значението на термина „големи данни“.

Таблица 1.1. Дефиниции на „големи данни“

Автор	Дефиниция
Gartner, 2011	Големите данни не може да бъдат събирани, управлявани и обработвани чрез конвенционални инструменти и похвати в рамките на разумно допустимото за потребителя време [35].
Gantz and Reinsel (IDC), 2011	Данни с голям обем и голямо разнообразие, които изискват ново поколение технологии и архитектури с цел извличане на стойност чрез улавяне, откриване и/или анализ с висока скорост [104].
Beyer and Laney, 2012	Информационни активи с голям обем, скорост и многообразие на източниците, които изискват ефективни и иновативни форми на обработка с оглед извличане на скрити смислени зависимости и вземане на информирани решения [35].

Автор	Дефиниция
McKinsey Global Institute, 2012	Масиви от данни, чиито размери надхвърлят възможностите за събиране, съхраняване, управление и анализ на традиционните софтуерни инструменти за работа с бази от данни [153].
Cooper and Mell, 2012	Данни, на които обемът, скоростта на придобиване или представянето им (разнообразие) ограничават възможностите за извършване на ефективен анализ, като използват традиционни релационни подходи [63].
IBM, 2012	Големите данни обединяват структурирани и неструктурирани данни (текст, сензорни данни, аудио, видео, клик стриймовете, логфайлове и др.), които се характеризират с огромен размер, хетерогенност и висока скорост на генериране и обработка. Целта на обработката и анализа на големите данни е получаването на полезно знание [128].
Dijcks, 2012	Големите данни имат 4 основни характеристики: обем, скорост, многообразие и стойност [82].
Intel, 2012	Огромен масив от сложни, неструктурирани данни [129].
Schroeck et al., 2012	Големите данни се характеризират с обем, разнообразие, скорост и истинност, които създават възможност за организацията да спечелят конкурентно предимство в днешния дигитализиран пазар [201].
Boyd and Crawford, 2012	Културен, технологичен и научен феномен, който се основава на взаимодействието между технология, анализ и митология [45].
Demchenko, 2013	Големите данни се характеризират с огромен обем, висока скорост и многообразие и изискват ефективни, новаторски форми на обработка за извличане на стойност и гарантиране на висока достоверност с цел подобряване на процесите на вземане на решения и контрол на процесите. Необходими са нови модели на данните, нови услуги и инструменти, които позволяват получаване и обработка на данни от различни източници (вкл. сензорни мрежи) и доставянето им в различни форми до различните потребители на информация и устройства [78].
Microsoft, 2013	Процес на прилагане на сериозна изчислителна мощност в машинното обучение и изкуствения интелект към масивни, често много сложни набори от информация [159].
Mayer-Schönberger and Cukier, 2013	Феномен, носещ три основни промени в начина на анализиране на информацията, която трансформира начина, по който разбираме и организираме обществото: 1. повече данни; 2. замърсени (непълни) данни; 3. зависимости, които не са следствие от причинно-следствени връзки [157].
Christozov and Toleva-Stoimenova, 2014	Масиви от данни, чийто размер е близък до горната граница по обем и сложност, които човек може да управлява с наличните към момента информационни технологии (ИТ) [53].

При дефинирането на термина „големи данни“ много от гореизброените автори се базират на основните им характеристики. Друг подход при дефинирането на големите данни е свързан с технологичните аспекти при обработката им. Той акцентира върху необходимостта от базови технологии, включващи инструменти и процеси, чрез които организациите да се справят с нарастващия обем и правилното използване на информацията. Например Microsoft определя големите данни като данни, които се нуждаят от „сериозна изчислителна мощност“ [159]. Някои автори се фокусират върху влиянието на големите данни в различни сфери на общественото развитие. [44, 45, 157]

В общия случай понятието „големи данни“ се свързва с огромни масиви от разнородни неструктурирани данни, получени от различни източници, които не може да се обработят чрез познатите технологии и похвати, а изискват създаването на нови, специализирани инструменти за извличане на скритата в тях полезна информация.

1.1.3. Проблеми и предизвикателства

Спецификата на големите данни води до възникването на следните важни проблеми:

- Стандартното търсене се използва за откриване на знания, за които вече има информация. За откриването на нови, неизвестни факти и зависимости патерните и други познати технологии не са подходящи.
- Инструментите, базирани на заявки, отнемат много време, защото за извличането на полезните знания от големите данни чрез подходящите, базирани на търсене, са необходими почти безкраен брой заявки.
- Статистическите методи до голяма степен са разработени за числови данни и не са подходящи за неструктурирани данни.
- Важна особеност на обработката на големите данни е необходимостта от холистичен, цялостен, многомерен поглед върху представената реалност, което предполага обработка на огромния информационен поток като цяло с цел получаване на пълноценно знание чрез провежданите анализи.

Поради това възниква необходимост от изграждане на нови инструменти и методи, които са специализирани за обработка и анализ на големи данни, както и на нови архитектури за тяхното съхранение и управление.

„Интернет на нещата“ (*Internet of Things, IoT*) през последните години се утвърди като един от основните нови източници на големи данни. През годините се присъединяват все повече устройства и машини, които генерират огромни количества данни от различни заво-

ди, изследователски станции, умни градове и др. Според Business Insider в световен мащаб до 2020 г. се очаква да има повече от 34 млрд. устройства, свързани към интернет. Информацията, събирана от разнородни интелигентни измервателни уреди, може да подпомогне подобряването на производствената, енергийната и т.н. ефективност на процесите. Затова търсенето на ефективни решения за събиране, съхраняване, анализиране и споделяне на генерираните големи данни в (реално) време е един от основните въпроси, свързани с управлението на устройствата в „интернет на нещата“.

Големи данни се получават също в производството, здравеопазването, образованието и други обществени сфери. Ежедневно обемът на съхраняваните в компютърните системи данни расте експоненциално и това създава много нови възможности.

Големите данни предполагат нещо повече от анализ на огромни обеми от информация. През последните години системите за работа с големи данни претърпяха значително развитие. Днес те се използват за извличане и разкриване на неочевидни зависимости, скрити корелации между на пръв поглед неизвестни и несвързани една с друга величини или за обобщаването им по нови начини, които са разбираеми и полезни за потребителите. За тази цел се прилагат интелигентен анализ на данни и текст (*Data Mining* и *Text Mining*), различни аналитични техники и методи за мрежов анализ, анализ на мултимедии и социални медии и др. Извлечените практически полезни закономерности и знания се използват за подпомагане на вземането на обосновани решения в различни области на човешката дейност.

Основните проблеми и предизвикателства пред изследователите са свързани с жизнения цикъл на данните и може да бъдат групирани в три основни категории в зависимост от особеностите на данните, процесите и управлението:



Фиг. 1.2. Жизнен цикъл на данните (адаптирана от Sivarajah [208])

Все още рамките, мащабът и обхватът на дейностите, свързани с големите данни, са в начален период на развитие. Своео адекватно решение трябва да намерят три категории предизвикателства.

Първото предизвикателство е свързано с характеристиките на събраните масиви от данни и извличането на полезна информация с добро качество, точност, достоверност и др.

Обработката, съхранението и трансферът на големи масиви от данни създават предпоставки за възникване на чисто технологични предизвикателства. По-нататък ще бъдат разгледани основните методи, техники и технологии за обработка и анализ на големи данни, състоянието и тенденциите им на развитие.

Често потребителите на услуги и устройства, генериращи данни, не са информирани, че ще бъдат включени в голямото количество данни и/или за какво тези данни може да се използват впоследствие. Третото ключово предизвикателство се отнася до някои етични аспекти при анализа на големи данни, които засягат поверителността и правните проблеми, свързани с авторското право и собствеността на данните. Въпросите, свързани с управлението на големи данни, са много важни, доколкото анализаторите работят с лична и чувствителна информация.

1.1.4. Съхранение и обработка на големи данни

Релационните бази данни преобладават при избора на модел за съхранение на финансови, производствени, лични и други видове данни. Релационните (*SQL*-ориентирани) бази съхраняват данните по предварително зададена структура, като таблици, подредени в редове (записи) и колони (атрибути/свойства/параметри). Между отделните данни и таблици е възможно да се създават връзки. Таблиците и връзките между тях образуват структура, която може да се представи като схема на различни нива на общност.

Днес наличието на огромни обеми неструктурирана информация и нуждата от висока скорост на достъпност и постоянно и автоматично разширяване налагат търсенето на нови решения за съхраняване на такива данни. Това изисква гъвкави бази данни, чието логическо и физическо съдържание може да се адаптират в синхрон с бързата еволюция на данните [121].

Подходящата платформа за анализ на данни трябва да включва следните области [31, 146]:

- способност за събиране на данни от различни формални източници в широк спектър от вътрешни и външни платформи;

- възможност за включване на различни типове данни от съществуващи бази данни до магазини за файлове (офисни и текстови документи), потоци от данни (например машинни данни от производствени линии, системи за сигурност на стради) и все повече изображение, глас и видео;

- способност за смесване по интелигентен начин на източници на информация така, че да може да се извърши цялостен анализ в по-широк набор от налична информация, а не само в един силов;

- възможност за извършване на анализ от технически изпитатели, използващи опростен интерфейс, а не ограничаване само до висококвалифицирани анализатори на данни;

- възможност различни хора да разглеждат резултатите по начин, който най-добре отговаря на техните нужди, вместо да използват предварително определени и структурирани отчети;

- способност за използване на потоци от процеси, които интелигентно да придвижват резултатите от анализа по веригата за вземане на решения;

- способност за постоянно проследяване на надеждността на съществуващите източници на данни.

За обработка и управление на данни от нерелационен тип (неструктурирани и полуструктурирани данни) се използват нерелационни бази данни *NoSQL (Not only SQL)*. С *NoSQL* бази данни се разрешават различни *SQL* ограничения, като:

- лесно мащабиране върху кластери от сървъри (хоризонтално мащабиране);

- поддржане на различни типове структури от данни;

- използване при разработка с гъвкави методологии (*agile development*).

Нерелационните бази данни, за разлика от релационните, не използват схеми и таблици за данните. Според модела на съхранение на данните се различават четири основни типа: документни бази, ключ-стойност бази/хранилища, колонни бази и мрежови бази [239, 240].

Документните бази данни съхраняват информация и свързаните с нея допълнителни данни в документ в *JSON*, *XML* или друг подходящ формат. Ключ-стойност базите са организирани в записи тип „ключ-стойност“, като колоната „стойност“ може да съдържа различни и сложни структури от данни. Колонните бази данни се базират на организацията на данните във фамилии от колони. Отново се

използва понятието „ключ“, но то може да сочи определена „фамилия“ колони в базата. Мрежовите бази данни са подобни на документните, но поддържат връзки между отделните обекти. В тези бази се използва идеята за „обект“ от данни, наричан „възел“ (node), който може да има свойства (properties) и връзки (edges) с други възли, формиращи ориентиран граф.

Популярни *NoSQL* системи са *MongoDB*, *Couchbase*, *Riak*, *Memcached*, *Redis*, *CouchDB*, *Hazelcast*, *Apache Cassandra* и *HBase* (продукти с отворен код) и *Google BigTable* и *Amazon Dynamo* (продукти със затворен код). От *Dynamo* и *BigTable* се заражда *Cassandra*, която се свързва с Facebook. От архитектурата на Google с GFS + *BigTable* се появява *Hadoop*, а от архитектурата на *Dynamo* – *Riak*. Съхранението на данни *ключ-стойност* еволюира, от една страна, до документния модел и базите данни *CouchDB*, *MongoDB* и по-късно *RavenDB*, а от друга – до кеш-ориентираните бази, като *Redis* и в по-малка степен – мрежовата база *Neo4j*. Таблица 1.2 съдържа класификация на нерелационните бази данни в зависимост от модела на съхранение на данните в тях.

Таблица 1.2. Модели и технологии за големи данни

Модел	Популярни <i>NoSQL</i> системи
Документни	<i>CouchDB</i> , <i>MongoDB</i> , <i>Apache Jackrabbit</i> , <i>RavenDB</i> , <i>Terrastore</i> , <i>Azure Cosmos</i>
Ключ-стойност	<i>Riak</i> , <i>Amazon Dynamo</i> , <i>Cassandra</i> , <i>BigTable</i> , <i>MemcacheDB</i> , <i>Redis</i> , <i>SimpleDB</i> , <i>Kyoto Cabinet</i> , <i>Tokyo Cabinet</i>
Колонни	<i>Maria</i> , <i>Apache HBase</i> , <i>BigTable</i> , <i>Mnesia</i> , <i>Hypertable</i>
Мрежови	<i>Neo4j</i> , <i>Titan</i> , <i>OrientDB</i> , <i>AllegroGraph</i> , <i>InfoGrid</i> , <i>DEX</i> , <i>FlockDB</i> , <i>InfiniteGraph</i>

Характерна черта на приложението на нерелационните бази данни са разпределеността и слабата свързаност между отделните компоненти. Поради това те са част от предложенията на повечето платформи за облачни изчисления. Нерелационните бази данни обикновено са с отворен код и възможност за работа в облак. Използват се при работа в реално време в уебсреда с големи обеми от данни. Този вид бази данни правят възможно съществуването на уебслужби, които не биха могли да съществуват по друг начин: *Hadoop* и *Cassandra* движат Facebook, Twitter и отделни услуги в Yahoo, SourceForge и Foursquare използват *MongoDB*. Основни услуги на Yahoo са *Sherpa*, на Google – *BigTable*, на Amazon – *Dynamo*, като всички гореизброени ползват в определени случаи и други *NoSQL* решения.

1.1.5. Обработка и анализ на големи данни

Според He, 2011 има четири критични изисквания при обработката на големи данни [119].

Първото изискване е свързано с **бързото зареждане на данните при изпълнението на заявките**. Turkaу et al. [223] твърдят, че приблизителните, но прогресивно подобряващи се резултати трябва да се предоставят на анализаторите своевременно. Според тях това може да трансформира начина, по който се извършва процесът на многостъпно откриване на знания, когато се въведе прогресивност в рамките на всяка стъпка.

Второто изискване се отнася до **бързината на обработка на заявки**. За да се задоволят изискванията за натоварване и изпълнение на заявките в реално време, времето за реакция е от решаващо значение. Съхраняването и обработката на междинните резултати трябва да може да се извършват с висока скорост при голямо количество на постъпващите заявки.

Третото изискване за обработка на големи данни е **високоефективното използване на пространството за съхранение**. Бързият растеж на броя на заявките изисква нарастващ капацитет за съхранение и изчислителна мощност. От своя страна ограниченото дисково пространство налага ефективно управление и съхранение на данните. Според Turkaу [223] дейностите, като селекция на данни, предварителна обработка, трансформация и извличане, често представляват повтарящи се процеси от типа „проба – грешка“. Те може да се ускорят значително чрез осигуряване на бърза обратна връзка за осъществяване на адекватно въздействие.

Четвъртото изискване е свързано с изискването за адаптивност при динамично развитие на натоварването в системата. Според He и Lee [119], тъй като наборите от големи данни се анализират от различни приложения и потребители за различни цели и по различни начини, основната система трябва да бъде силно адаптивна към промените при обработката на данните. Авторите откриват, че изчисляването на резултатите от осъществените промени трябва да се извършва по прогресивен начин, при осигуряване на бързо връщане към първото сближаване на резултатите и допускане на итеративни подобрения до сближаването с крайния резултат. Осигуряването на взаимодействие на потребителя с междинните резултати позволява рано откриване на погрешни или неоптимални решения и тяхната бърза оценка [223].

Анализът на данните представлява процес на прилагане на алго-

ритми, за да се анализират набори от данни и да се извлекат полезни и непознати модели, взаимоотношения и информация [22].

Според Гаууад изборът на алгоритъм за извличане на данни стои в основата на процеса на откриване на знания [100]. Разработени са няколко онтологии за избор на алгоритъм при извличането на знания от данните, за да се предостави адекватна помощ в този подбор. Например *OntoDM* е онтология на най-високо ниво за концепции за извличане на данни. Тя описва основни единици, целящи да обхванат цялата същност на *Data Mining*, докато онтологията *EXPO* е фокусирана върху моделиране на научни експерименти [176].

По-пълна онтология е т.нар. *DMOP* [123, 124], която описва не само алгоритмите за обучение (включително техните вътрешни механизми и модели), но и работните процеси. Голям брой оператори за извличане на данни са описани в онтологията *KD* [241] и *eProPlan* [135].

По отношение на работните процеси за извличане на данни онтологията *KDDONTO* [83] цели както откриване на подходящи *KD* алгоритми, така и описание на работните потоци на *KD* процесите. Основно се фокусира върху концепции, свързани с входовете и изходите на алгоритмите, както и на всякакви предварителни условия и следусловия за тяхното използване.

В свое изследване Vanschoren et al. предлагат специфична онтология за описване на експерименти с машинно обучение по стандартизиран начин за подпомагане на съвместния подход при анализа на алгоритмите за обучение [40, 226].

Съществуват проекти, които позволяват на научната общност да допринесе със своите експерименти за подобряване на процеса на откриване на знания [95]. Подобна база данни за експерименти с машинно обучение е разработена от University of Leuven и предлага уеб-инструмент за съхранение на експериментите, извършени в база данни, и възможност за заявка [40]. Проектът e-LICO, финансиран от Седма рамкова програма, има за цел да разработи управляван от знания помощник за извличане на данни (*Knowledge-Driven Data Mining Assistant*), който разчита на онтологията при планиране на процеса *Data Mining* [123, 124].

Съществуват разнообразни методи за анализ на огромни масиви от данни с цел извличане на полезна информация и подобряване на процесите на вземане на решения. Те може да се разделят на две основни групи, съответно заимствани от статистиката и изкуствения интелект:

- **методи на математическата и приложна статистика** за обработка и анализ на масови съвкупности от данни, като дескриптивен анализ, анализ на връзките (корелационен анализ и регресионен анализ, факторен анализ, дисперсионен анализ), многомерен статистически анализ, анализ на динамични редове и прогнозиране;

- **методи, основаващи се на теорията на изкуствения интелект и машинното обучение, и евристични методи**, чрез които може да бъде изследвана връзката между състоянието на съвкупността от входни данни или предпоставки и изходния резултат. Обикновено те позволяват да се моделира поведението на реални технически, технологични или икономически процеси, без да се анализира в детайли тяхната структура. В тази група се включват изкуствени невронни мрежи, еволюционно програмиране, генетични алгоритми, асоциативна памет, разпита логика и изчисления, дървета на решенията и др.

В Таблица 1.3 е предложен списък с аналитични методи, използвани при обработката и анализа на големи данни [50, 57, 153].

Таблица 1.3. Методи за анализ на големи данни

Метод	Описание
A/B тест (A/B testing)	Сравняване на контролен обект с много тестови обекти с цел неговото подобряване.
Извличане на асоциативни правила (Associated Rule Mining)	Откриване на интересни взаимовръзки между променливите, аналозите и прототипите на поведение в големи бази данни.
Класификация (Classification)	Изграждане на модели, описващи съществуващи класове от обекти, с цел използването им за определяне на класа на обекти с неизвестен клас.
Клъстерен анализ (Cluster Analysis)	Разделяне на данните на групи (клъстери) чрез прилагане на принципа за максимално сходство между елементите на един клъстер и минимално сходство между елементите на различни клъстери.
Краудсорсинг (Crowdsourcing)	Генериране на идеи/иновации чрез сътрудничеството на голяма група хора.
Обобщаване и интегриране на данни (Data Fusion and Integration)	Обобщаване на набори от данни с цел последващото им анализиране като едно цяло, а не независимо, за да се създаде по-общ поглед и да се разкрият скрити модели и корелации.
Извличане на знания от данни (Data Mining)	Интелигентен анализ на данни с цел разкриване на скрити модели и закономерности, повтарящи се образци, тенденции за предвиждане на бъдещото изменение на параметрите.

Метод	Описание
Ансамбловото обучение (Ensemble Learning)	Техника, използвана за целите на машинното обучение (при обучение с учител), включваща множество предикативни модели, благодарение на което постига високо качество на изведените прогнози.
Генетичен алгоритъм (Genetic Algorithm)	Евристичен алгоритъм за търсене, който се основава на принципите на естествената еволюция. Използва се за решаване на оптимизационни задачи и моделиране чрез случаен подбор, комбиниране и вариации на целеви параметри и използване на механизми, аналогични на естествения подбор.
Машинно обучение (Machine Learning)	Направление от областта на изкуствения интелект, насочено към създаването на алгоритми за самообучение на базата на емпирични данни.
Машинна обработка на текст на естествен език (Natural Language Processing)	Вид машинно обучение за разпознаване и обработка на естествен език, заимствано от информатиката и лингвистиката.
Невронни мрежи (Neural Networks)	Имитират човешкия мозък за разпознаване на образи и реч, използват обучение със и без учител.
Мрежов анализ (Network Analysis)	Група методи за анализ на връзки между възли в мрежи (графи). Прилагат се към данни от социалните мрежи, като позволяват да се анализират връзки между отделни потребители, компании, общности и др.
Предиктивно моделиране (Predictive Modelling)	Метод, който позволява с помощта на статистически и математически техники да се създаде модел на предварително зададен вероятен сценарий за развитието на дадени събития.
Регресионен анализ (Regression)	Статистическа техника, използвана за прогнозиране, която определя причинно-следствени връзки между две или повече променливи.
Анализ на чувствата (Sentiment Analysis)	Методи за идентифициране и оценка на настроенията на потребителите (емоционална наситеност, мнения и нагласи и др.), базирани на технологии за разпознаване на естествен език.
Пространствен анализ (Spatial Analysis)	Използва се за изследване на географски данни – топология на местност, географски координати, геометрия на обекти и др.
Статистика (Statistics)	Събиране, организиране, анализ и интерпретация на данни от проучвания или експерименти.
Обучение с учител (Supervised Learning)	Идентифициране на скрити в данните модели въз основа на предварително зададени образци (обучаваща съвкупност), описващи наблюдаваното явление или обект.

Метод	Описание
Симулация (Simulation)	Симулиране на поведението на някои сложни системи с цел прогнозиране и планиране на изходните променливи и измерване на резултатите.
Обучение без учител (Unsupervised Learning)	Идентифициране на скрити в данните модели без априорни знания. Пример за обучение без учител е кластерният анализ.
Визуализация (Visualization)	Методи за графично представяне на резултатите от анализа на големи данни с помощта на графики, диаграми, изображения и др. с цел опростяване на интерпретацията и по-лесно разбиране на получените резултати.

С цел извличане на полезни знания от съдържанието на документите в интернет, напоследък се развива *анализът на чувствата/нагласите* (*Sentiment analysis*). Често се нарича още *извличане на чувства* (*Sentiment mining*) или *извличане на мнения* (*Opinion mining*) [39]. Базира се на технологиите за обработка на естествен език, като се фокусира върху определянето на схващанията, възгледите, идеите и разбиранията по дадена тема с цел да извлече атрибути, характеристики на обекта и въз основа на тях да определи дали коментарите са положителни (възторг, радост, щастие, обич и др.), отрицателни (тъга, яд, страх и др.), или неутрални (изненада, безразличие, безпокойство и др.). Една от основните задачи при анализа на чувствата е класифицирането на документи или изречения в един от двата класа – обективни или субективни. Анализират се само субективните, а обективните, които представляват фактологична информация, не са предмет на анализ. Анализът на чувствата се извършва на три нива – документи, изречения и фрази/думи. При анализа на ниво документ се изследва общото становище, представено в документа – положително, отрицателно или неутрално. Този анализ спомага за установяване на цялостната полярност на текста, но чрез него не може да се оценят различните емоции около отделни аспекти на текста. В анализа на ниво изречение се изследва полярността на всяко изречение, а на ниво дума се изследват специфични думи, които са носители на емоции и настроения. В литературата често се използват групиране на думите в речници и тяхната автоматична класификация, извличане на ключови думи и идентифициране на теми, и др. [120, 127, 190, 216].

Според Ravi, Ravi, 2015 анализът на чувствата се занимава с класифициране на текста според субективността му и изразените в него настроения, с изследване на полезността на мненията, с откриване на спам текстове, със създаване на речници и лексикони и извличане на характерни черти и аспекти [193].

Безспорно много големи възможности за анализ на големи данни притежават *усъвършенствана визуализация на данни (Advanced Data Visualization - ADV)* и *визуално откриване (Visual discovery)* [200]. Представянето на информация така, че хората да могат да я „консумират“ ефективно, е ключово предизвикателство, което трябва да бъде изпълнено, за да могат отговорните лица да анализират правилно данните по начин, който да доведе до вземане на информирани решения и конкретни действия [153].

ADV се очертава като мощна техника за откриване на знания от данните, която комбинира методи за анализ на данни с интерактивна визуализация, за да даде възможност за изчерпателно проучване на данните. Това е изследователски подход, базиран на данни, който се вписва добре в ситуации, при които анализаторите нямат достатъчно познания за данните [205].

С генерирането на все повече големи данни се появява все по-голямо търсене на *ADV* решения в различни области на приложение [243]. Наред с големината и сложността на големите данни е необходимо осигуряването на интуитивно визуално представяне и взаимодействие, за да се улеснят възприемането и разсъжденията на анализатора [205].

ADV може да даде възможност за по-бърз анализ, по-добро вземане на решения и по-ефективно представяне и разбиране на резултатите чрез предоставяне на интерактивни статистически графики и интерфейс *point-and-click* [48]. *ADV* е подходящ за работа с големи данни, тъй като може да мащабира визуализациите си, представяйки хиляди или милиони точки за данни (за разлика от стандартните диаграми). Освен това той може да работи с различни типове данни и структури от данни и поддържа интерфейси към всички водещи източници на данни.

1.2. НАУКА ЗА ДАННИТЕ

1.2.1. Основни дефиниции

За първи път дефиниция за науката за данните се появява в книгата *Concise Survey of Computer Methods*, в която авторът я определя като „наука, която се занимава с данните от момента на тяхното създаване до предоставянето на връзката с това, което представят, на други области и науки“ [166].

John W. Tukey има голям принос в развитието на обработката и

анализа на данни. За първи път той дефинира понятието „изследователски анализ на данни“ (*Exploratory Data Analysis – EDA*) [222]. През последните години анализът на данни се разви и се появиха научни направления с различни наименования, като *Data analysis*, *Data Analytics*, *Advanced Analytics*, *Big Data Analytics*, *Deep Analytics*. Утвърдиха се области като *Data Mining*, *Knowledge Discovery* и *Machine Learning*, които са свързани с интелигентен анализ на данните и разкриване на закономерностите, скрити в тях. С напредъка на ИКТ и изчислителната мощ на компютрите се развиха и подобриха механизмите за управление на данните (*Data Management*), за съхранение, извличане и възстановяване на данни, за изпълнение на заявки, обработка на трансакции и разширен анализ на данните. Науката за данните може да се разглежда и като резултат от естествената еволюция на всички направления, пряко или косвено свързани с анализа на данни.

Науката за данните стана популярна през последното десетилетие с разцвета на много големи интернет корпорации, като Yahoo, Google, LinkedIn, Facebook и Amazon, а също и стартиращи компании, базирани на данни, като Palantir, Everstring, Climate Corporation и Stitch Fix.

В литературата се разглеждат два подхода за дефиниране на науката за данните – „от горе надолу“ и „от долу нагоре“. От една страна, ако се гледа от горе надолу, науката за данните е изследователско поле, което изучава механизми и подходи, необходими за генериране на полезно знание и получаване на бизнес стойност от данните [148]. От друга страна, ако се възприеме възгледът „от долу нагоре“, науката за данните е интердисциплинарна изследователска област [214], която използва нов, холистичен начин на експлоатация на данните, гледайки отвъд аспекти като съхранение или достъп например. Според Christozov et al. науката за данните е пример за трансдисциплинарна област, която се характеризира с постигане на синергия чрез набор от разнообразни области на знанието – от технологични новости и обработка на големи данни чрез интензивно използване на техники за анализ на данни до различни, определени от конкретната предметна област, начини за представяне и визуализиране на резултатите в компактна и информативна форма [56].

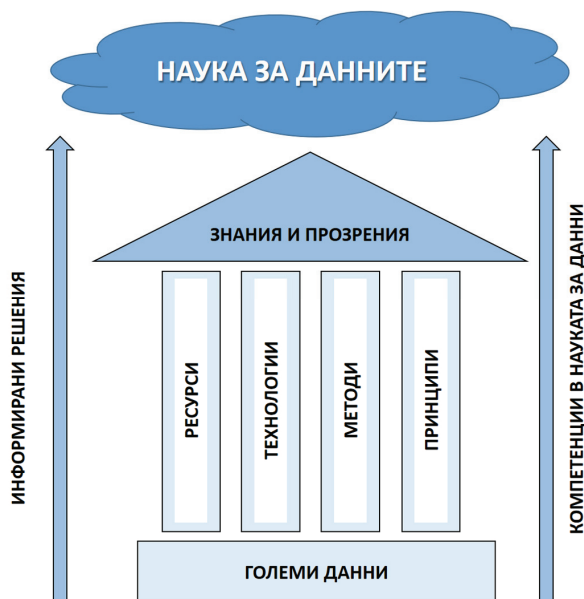
В търсене на единна дефиниция за науката за данните Parks стига до заключението, че обединяващото във всички съществуващи дефиниции са ключови фрази като **„включва комбинация от статистика, математика и програмиране“**, както и **„с цел извличане на смислена информация от данните“**[177].

В следващата Таблица 1.4 е дадено хронологично представяне на различни определения за науката за данни, получени в резултат от проучване на литературата по въпроса.

Таблица 1.4. Дефиниции за науката за данните

Автор	Дефиниция
Provost and Fawcett, 2013	Включва множество основни принципи, които подпомагат и управляват извличането на информация и знания от данни [187,188].
Dhar, 2013	Област, която се фокусира върху събирането, извличането и преобразуването на данните в знания [80].
Shum et al., 2013	Науката за данни е комбинация от статистика, компютърни науки и информационен дизайн [203].
Andrejevic, 2014	Предлага нови начини за използване на данни, необходими за осъществяване на прогноза и формиране на решения, влияещи върху всички сектори – от здравеопазване до градско планиране, финансово планиране, скрининг за работа, прием на образование и др. [27, 28].
Diggle, 2015	Включва не само статистика, но и информатика (хардуерно и софтуерно инженерство) [84].
Parks, 2017	Комбинация от техники, знания и умения, прилагани върху данните, за да се идентифицират скрити знания, които може да се използват за насочване на екипи при вземане на важни решения [177].
Сао, 2017	Нова интердисциплинарна област, която включва статистика, информатика, изчислителна техника, комуникация, мениджмънт и социология с цел трансформиране на данните в прозрения и решения, като се следва схващането за йерархията „данни – знание – мъдрост“ [47].
Turkay et al., 2018	Процес на многоетапно откриване на знания, в който от сурови, често нечисти колекции се извличат полезни знания в определен контекст [223].
Yan and Davis, 2019	Предоставя принципи, методология и насоки, служещи при анализ на данни за: (1) инструмент (инструменти за визуализация, събиране на данни или проучване), (2) стойност (търговска или научна) или (3) знание (скрити обективни практически полезни зависимости [238].

На Фиг. 1.3 е предложена обобщаваща диаграма, представяща същността на науката за данните.



Фиг. 1.3. Наука за данните – обобщаваща рамка на дефинициите

Обект на науката за данните са наличните големи данни.

Предмет на науката за данните са (1) информираното вземане на решения (от страна на бизнес мениджърите) и (2) развиването на съответните компетентности от страна на специалистите по данни.

Търсен резултат в следствие на приложените методи, техники, принципи и налични ресурси за работа с данни е формирането на знание и прозрение от наличните данни.

В следващата точка ще разгледаме концепциите за мулти-, интер- и трансдисциплинарност на науката за данните с оглед на широкия спектър от области, които обхваща, и необходимостта от изграждане на многообразни умения у специалистите по данни.

1.2.2. Мулти-, интер- и трансдисциплинарност на науката за данните

Вследствие на повсеместното използване на ИКТ и развитието на обществото, на повечето науки в днешно време им се налага да излязат от своите граници и да използват нови методи и технологии, позволяващи изследвания и обработка на данни. Интер-, мулти- и трансдисциплинарните изследвания се превръщат в обичайна практика.

Днес е често срещана практика мултидисциплинарни екипи, включващи експерти от две или повече дисциплини, да работят заедно върху решаването на общ проблем. Всеки участник използва теории, методи и техники от своята област. По този начин се включват множество гледни точки и разнообразен инструментариум, което повишава ефективността на работата. В процеса на прилагането на специфично дисциплинарни методи към различни научни области те се имплицират в мултидисциплинарни методи.

От друга страна, интердисциплинарният подход обединява две или повече утвърдени дисциплини, които си взаимодействат и по този начин се разширяват предимствата, които носи всяка от тях. Интердисциплинарността интегрира информация, данни, техники, инструменти, перспективи, теории от две или повече дисциплини, за да се постигне фундаментално разбиране или да се решат проблеми, чиито решения са извън обхвата на една-единствена дисциплина [26]. Преведено буквално, „мулти-“ означава наличие на множество, а „интер-“ означава насоченост навътре, взаимодействие. С цел анализране и разбиране на сложността на обекта на изследване или решаване на теоретичен или практически проблем това взаимодействие трябва да доведе до интегриране на различните допълващи се гледни точки сред партньорите, участващи в процеса на изследване.

Науката за данните представлява мултидисциплинарна област, обхващаща инструменти, методи и системи, като статистически анализ на данни, прилаган върху големи обеми от данни, с цел да се разберат причинно-следствени зависимости, служещи при вземане на решения [86, 140, 167, 187, 188, 232]. Напредъкът на ИКТ дава възможности за развитие и усъвършенстване на съществуващи, както и за създаване на нови методи, техники и технологии, заимствани от статистиката и изкуствения интелект за интелигентна обработка на (големи) данни.

Интердисциплинарността на науката за данните се определя от факта, че тя не може да се побере в традиционните класификации на единична дисциплина, а обхваща области като статистика, информатика, компютърни технологии, комуникация, управление, социология, психология и др., които дават свои критерии, но заедно изграждат и утвърждават нова специфична област.

Parks разглежда интердисциплинарността на науката за данните от гледна точка на професионалния профил на специалистите, изследващи данните, и „интердисциплинарните знания“, които трябва да притежават [177]. Според авторката към тази категория принадле-

Трите концепции си приличат по общата идея за единство, взаимни отношения, интегриране на научните дисциплини, трансфер на методи между науките и т.н., която споделят. Всички те се отнасят до значението на нецентрализирането на мисълта в една наука, но чрез интегрирането на няколко науки за постигане на успех в изследването на сложни многомерни процеси и явления.

Според Christozov науката за данните има трансдисциплинарен характер и това трябва да е водещ принцип при обучението в тази област и изграждането на компетенциите, необходими на специалистите, които работят с данни. Необходимо е у тях да се формират холистичен поглед върху проблемите и умения за търсене на взаимовръзки и зависимости, както и за по-широка перспектива при решаване на проблемите [56].



В литературата съществуват множество разнообразни схващания за обхвата на науката за данните и професионалния профил на специалистите по данни. Често обсъждана тема е голямото разнообразие от специалности, които трябва да притежават анализаторите, и дали въобще съществуват такива „еднорози“ [106]. В следващата Таблица 1.5 са поместени популярни диаграми на Вен, представящи различни гледни точки за включените области, които показват мулти-, интер- и трансдисциплинарния характер на науката за данните.


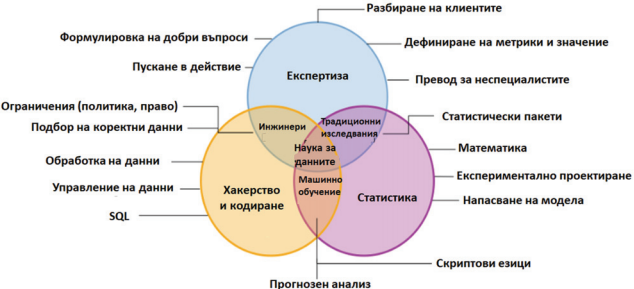

Таблица 1.5. Диаграми на Вен за обхвата на науката за данните



Автор	Диаграма
Tierney, 2012 [220]	
Harris, 2013 [116, 117]	

Автор	Диаграма
Matter, 2013 [156]	 <p>A Venn diagram with three overlapping circles. The top circle is light blue and labeled "ОБЩЕСТВЕНИ НАУКИ". The bottom-left circle is light green and labeled "КОЛИЧЕСТВЕНИ МЕТОДИ". The bottom-right circle is light red and labeled "КОМПЮТЪРНИ НАУКИ". The intersection of the top and bottom-left circles is labeled "Традиционни емпирични изследвания". The intersection of the top and bottom-right circles is labeled "Изчисления, базирани на данни". The intersection of the bottom-left and bottom-right circles is labeled "Умения в областта на науката за данни".</p>
Ayankoya et al., 2014 [29]	 <p>A Venn diagram with three overlapping circles. The top-left circle is yellow and labeled "Бизнес разумеаване, програмиране и визуализация на данни". The top-right circle is light blue and labeled "Статистически и аналитични техники". The bottom circle is pink and labeled "Знания в областта". The intersection of the yellow and light blue circles is labeled "Бизнес анализи". The intersection of the yellow and pink circles is labeled "Съхранение и управление на данни". The intersection of the light blue and pink circles is labeled "Емпирични изследвания". The central intersection of all three circles is labeled "Data Science".</p>

Автор	Диаграма
van der Aalst, 2014 [224]	
Geringer, 2014 [106]	

Автор	Диаграма
Palmer, 2015 [175]	 <p>ОПИТ В ОБЛАСТТА</p> <p>Статистически изследвания</p> <p>Обработка на данни</p> <p>НАУКА ЗА ДАННИТЕ</p> <p>МАТЕМАТИКА</p> <p>Машинно обучение</p> <p>КОМПЮТЪРНИ НАУКИ</p>
Eubanks, 2016 [96]	 <p>Разбиране на клиентите</p> <p>Дефиниране на метрики и значение</p> <p>Експертиза</p> <p>Превод за неспециалистите</p> <p>Формулировка на добри въпроси</p> <p>Пускане в действие</p> <p>Статистически пакети</p> <p>Математика</p> <p>Ограничения (политика, право)</p> <p>Подбор на коректни данни</p> <p>Инженери</p> <p>Традиционни изследвания</p> <p>Наука за данните</p> <p>Машинно обучение</p> <p>Статистика</p> <p>Обработка на данни</p> <p>Управление на данни</p> <p>Хакерство и кодиране</p> <p>SQL</p> <p>Експериментално проектиране</p> <p>Напасване на модела</p> <p>Скриптов езици</p> <p>Прогнозен анализ</p>
Silver, 2017 [207]	 <p>Опасна зона, когато институциите са в нужда</p> <p>АВТОМАТИЗАЦИЯ</p> <p>ВАЛИДНОСТ</p> <p>Хакерски умения</p> <p>Многовариантна статистика</p> <p>Машинно обучение</p> <p>Наука за данните</p> <p>Уни- и бивариантна статистика</p> <p>Традиционен софтуер</p> <p>Традиционни изследвания</p> <p>Експертиза в областта</p> <p>ИНТУИЦИЯ</p> <p>Опасна зона, когато е необходима валидация</p> <p>Опасна зона, когато са необходими многовариантни техники</p>

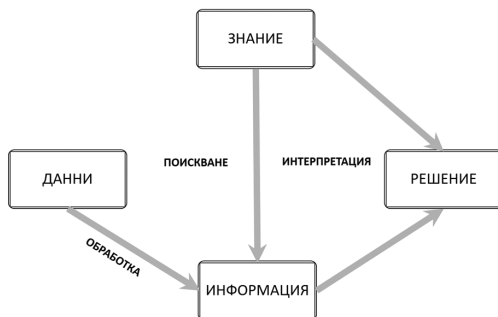
1.2.3. Процес на информационно осигуряване

Днес се наблюдава нарастване на стойността на информацията като продукт от гледна точка на информационното осигуряване. Извлечените скрити данни и информация се разглеждат като краен продукт, който е осъзнат като изключително ценен актив за организациите и изследователите и повишава тяхната конкурентоспособност.

Данните (от лат. *datum* – давам) са все още първични и сурови, примитивни единици, получени в резултат на измерване, експерименти, наблюдения, аритметични и логически операции и др. Те може да се представят чрез числови стойности, текст, графика, изображения, звук, видео и др.

Набирането на данни и създаването на бази данни водят до създаване на ефективни механизми за тяхното съхранение и управление. С развитието на ИКТ се създават възможности за по-задълбоченото им аналитично изследване и обобщаване с оглед на рационалното и ефективното им използване.

Данните придобиват смисъл в даден контекст след тяхното интерпретиране и разбиране и се превръщат в полезна **информация**, когато бъдат подбрани, организирани и обработени за определена цел. Информацията (от лат. *informatio* – разяснение, изложение, осведоменост) се представя чрез образци, схеми, асоциации, свързващи събраните данни. Информацията от своя страна подлежи на анализ, за да се открият връзки, зависимости, закономерности, които трябва да бъдат приети за достоверни, за да се превърнат в **знание**, което да мотивира и определя поведението на потребителя. Знанието е разбирането за съдържанието, което информацията носи. То е субективно и се основава на опита на отделния индивид. Използва се за вземане на решения при планирането и управлението на бизнес процесите.



Фиг. 1.5. Данни – информация – знание

Информацията и знанието се превръщат в най-важните и полезни ресурси на стопанската дейност и в ключови фактори за ефективно управление (Фиг. 1.5).

Специфичните управленски решения са краткосрочни решения, които са свързани с оперативната дейност на предприятието, нямат рутинен характер и са силно разнородни. Те включват например решенията за добавяне/закриване на оперативни сегменти, за производство/доставка на компоненти, ценовите решения във връзка с еднократни специални поръчки и др. Днес процесите на вземане на решение в организациите във всичките им етапи се подпомагат от информационни системи. В Таблица 1.6 е показано развитието на анализа на данни и извличането на знание за вземане на решения при планирането и управлението на бизнес процесите [72, 73].

Таблица 1.6. Еволюция на анализа на данни

Наименование	Период	Основни характеристики
Подпомагане на вземането на решения (Decision support)	1970 – 1985	Анализ на структурирани данни за подпомагане на процесите на вземане на решения.
Подпомагане на вземането на решения на изпълнителния директор (Executive support)	1980 – 1990	Анализ на данни за подпомагане на процесите на вземане на решения на ниво изпълнителен директор.
Системи за оперативна аналитична обработка (OLAP)	1990 – 2000	Аналитична обработка в реално време на многомерни бази данни.
Бизнес интелигентност (Business Intelligence)	1989 – 2005	Приложения и методологии, използвани за преобразуването на данните в информация и знание с цел подпомагане на вземането на управленски решения.
Анализ на данни (Analytic)	2005 – 2010	Анализ на данни чрез методите на статистическото и математическото моделиране с цел подпомагане на процесите на вземане на решение.
Анализ на големи данни (Big Data Analytics)	2010 – досега	Анализ на структурирани и неструктурирани големи масиви от данни, които се генерират и обработват в реално време.

Таблица 1.7 показва хронологичното развитие на технологиите за работа с данни през последните години.

Таблица 1.7. Развитие на технологиите за работа с данни

Ниво	Бизнес въпрос	Технология	Характеристики
Събиране на данни (1960)	Колко е средната печалба през последните пет години?	Компютри	Доставка на статични данни за минали събития.
Достъп до данни (1980)	Колко бройки са били продадени през месец март миналата година в САЩ?	Реляционни бази данни, SQL	Динамично доставяне на данни до ниво запис.
Навигационни данни (1990)	Колко бройки са били продадени през месец март в САЩ? Да се покаже до дълбочина Бостън.	OLAP, складове от данни	Динамично доставяне на данни на различни нива.
Извличане на знания от данни (2000)	Какво ще стане с продажбите в Бостън през следващия месец? Защо?	Алгоритми на високо ниво, многопроцесорни компютри, масивни центрове за данни (<i>Data Mining, Data Warehousing, Multimedia databases, Web databases</i>)	Проактивно получаване на информация за перспективите.

Наличието на огромни неизползвани масиви от данни днес дава нови възможности за тяхната обработка и анализ с цел разкриване на неочевидни зависимости, скрити корелации между на пръв поглед неизвестни и несвързани една с друга величини или за обобщаването и визуализацията им по нови начини, които са разбираеми и полезни за потребителите. Използват се интелигентен анализ на данни и текст, различни аналитични техники и методи за мрежов анализ, анализ на мултимедии и социални медии и др. Изведените от данните зависимости и обобщения се наричат **модели** или **шаблони** и се представят чрез уравнения, правила, кластери, графи, дървета и т.н. Получаването на дадена информация и полезно знание относно връзките и зависимостите между параметрите на управлявания процес е много важен процес, който предшества процеса на вземане на решение за извършване на конкретно действие. В този смисъл науката за данните свързва средствата и технологиите за събиране, съхране-

ние и достъп до данните, от една страна, и процесите, моделите и алгоритмите за вземане на решение, от друга. Затова специалистите по данни играят ключова роля в процесите на вземане на информирани решения в различни области на човешката дейност.

1.1.1. Основни приложения

Днес развитието на ИКТ дава възможност за съхранение и управление на големи обеми от данни. Науката за данните намира приложение в редица области за анализ и интелигентна обработка на събраните данни.

Образование. Съвременните университети представляват организации, в които се събират и съхраняват огромни обеми от данни, свързани със студентите, с методите и управлението на образователния процес, както и с цялостното администриране на дейността. В институциите, които предлагат традиционно обучение, данни се събират при приемането на нови студенти, при организирането и осъществяването на обучението, за целите на управлението и др. Анализването на тези уникални типове данни чрез прилагането на аналитични методи в областта на образованието води до появата на едно сравнително ново направление, зародило се през 2005 г. – извличане на знания от данни в сферата на образованието (*Educational Data Mining, EDM*) [199].

Производство. Големи данни се генерират във всеки сегмент на производството. Съответно анализът на големи данни може да се използва за подпомагане на процесите на вземане на решения, трансформация и иновиране на почти всички важни процеси в организациите, като маркетинг, идентифициране на най-добрите доставчици, наблюдение на работата на оборудването, поддръжка и предотвратяване на инциденти или прекъсвания, анализ и прогноза на потребителското поведение, разработване на нови продукти и др. [231].

Здравеопазване. [152, 173] Напоследък се използват **клинични интелигентни системи**, които за разлика от бизнес интелигентните системи (БИС) включват множество неструктурирани данни с различен формат и размер, като аудиозаписи, магнитен резонанс, компютърна томография и други диагностични изображения, електрокардиограми и т.н. Клиничната интелигентност изисква набор от аналитични методи и процеси за извличане и трансформиране на суровите данни в нови клинични открития и познания, които влияят върху вземането на решения в здравния сектор [118].

Банки и финанси. Основните предизвикателства в банковия

сектор са свързани с предотвратяването на измами в „онлайн банкирането“, прането на пари и анализа на риска. Сред най-важните рискови фактори, засягащи банките, са промените на лихвените проценти (лихвен риск), валутните курсове (валутен риск), цените на акциите (риск за цените на акциите), както и кредитния риск, свързан с идентификация на неплатежоспособни клиенти. Процесът по управление на риска представлява непрекъснато разглеждане на предприетите политики, следене и анализиране както на пазара, така и на клиентите.

Околна среда. Изследва се въздействието на производствената и преработващата промишленост върху околната среда [125, 133]. Освен това се следят климатичните промени и чрез различни технологии за извличане на данни се прилагат мерки за превенция и откриване на аномалии [98, 152].

Освен гореизброените съществуват още множество приложения в различни стопански сфери, например за енергийно управление на сгради [164], динамично ценообразуване и управление на доходността и др.

1.3. ДИГИТАЛНО НЕРАВЕНСТВО В КОНТЕКСТА НА ГОЛЕМИТЕ ДАННИ

1.3.1. Грамотност за работа с големи данни

Големите данни, които ежедневно се генерират от хора и машини, представляват голямо предизвикателство в съвременното дигитално пространство с оглед на ефективното им използване. За получаването на полезни знания от тях е необходима съответната грамотност, която притежават само специалистите по данни.

Стандартите за грамотност също се промениха значително през последните години. Общата грамотност включва способности за четене, писане и смятане, но днес те не са достатъчни за работа с компютърните информационни системи, които изискват и специални компютърни умения. В литературата съществуват множество дефиниции за цифрова, компютърна и ИКТ грамотност, а също и за информационна, медийна, интернет грамотност и т.н. Големите данни изискват допълнителни умения, които представляват и нов тип грамотност. Всички тези различни грамотности се припокриват и надграждат, а често цифровата грамотност се разглежда като събирателно понятие.

Цифрова грамотност представлява способността на потребителите да използват дигитални инструменти при осъществяването на

идентификация, достъп, управление, оценка на дигитални ресурси и общуване в контекста на конкретни житейски ситуации [108]. Цифровата грамотност предполага умения и компетентности за работа с компютърни технологии и интернет, като включва:

- базови дигитални умения (работа с компютър);
- умения, позволяващи търсене на информация в интернет;
- стратегически умения (например заплащане по електронен път).

Някои изследователи на цифровата грамотност определят четири основни компетентности, които предполагат усвояването на основни умения в областта на компютърната грамотност – ИКТ грамотност, интернет грамотност, медийна грамотност и информационна грамотност, които са тясно свързани помежду си [24].

Според модела на Eshet-Alkalai цифровата грамотност на потребителя е комбинация от технически, познавателни и социални умения [93, 94]. Тя представлява множество от пет грамотности: фотовизуална, репродуктивна, хипермедийна, информационна и социално-емоционална.

Фотовизуалната грамотност включва способности да се „чете“, тълкува и разбира информацията, представена в живописни или графични изображения, а също и способност информацията да се преобразува в картини, графики и др.

Репродуктивната грамотност се свързва със способности за интегриране на разнородни независими информационни ресурси в нова информация, която е смислена, достоверна, творческа интерпретация. Например такава грамотност се изисква при писането на академични статии и реферати.

Хипермедийната грамотност включва способности и умения за използване на World Wide Web, намиране, обработка и тълкуване на информацията в хипертекстова нелинейна форма. През последните години се използва понятието **грамотност за работа със социални медии и социални мрежи** [18]. Уеббазираните социални мрежи са интерактивни системи за информиране и споделяне на знание, които поставят нови изисквания за характеристиките на техните потребители.

Информационната грамотност не е понятие, което се появява през цифровата ера. Необходимост от информационна грамотност на потребителите на информация е съществувала още преди появата на компютрите и интернет. Този вид грамотност е основен елемент на процеса на извличане и обработване на информация и на процеса на учене. Информационната грамотност се основава на персонални познавателни способности, аналитично и критично мислене относно

качеството на добитата информация. Тя играе ролята на филтър при отделянето на значимата, релевантна информация относно поставения проблем от незначимата. Днес информационната грамотност е от съществено значение с оглед на намирането на търсената информация в огромното количество достъпна информация. Информационно грамотните потребители притежават необходимите умения за търсене и оценка на информацията от разнообразни източници (компютър, книга, медия, филм или друг вид ресурси), което е основен фактор за ефективното ѝ използване в определен контекст.

Информационната грамотност включва следните способности за работа с информация:

- осъзнаване кога възниква нуждата от информация;
- идентифициране на информацията, която е свързана с проблема;
- намиране на необходимата информация;
- оценяване на намерената информация;
- организиране на намерената информация;
- ефективно използване на информацията във връзка с поставения проблем;
- прилагане на информацията за създаване и комуникиране на знание.

Социално-емоционалната грамотност включва способности за осъществяване на ефективна виртуална комуникация, например социални мрежи, чатове, дискуссионни групи и др. Този вид грамотност е най-труден за дефиниране, защото е комплекс от социални аспекти, нагласи и неписани правила за комуникация в интернет. Тя изисква от потребителите високи нива на информационна и хипермедийна грамотност.

В модела на Eshet-Alkalai уменията, попадащи в обхвата на цифровата грамотност, прерастват от базови по отношение на достъп, информираност и обучение до творческа, критична и аналитична грамотност у потребителите [93, 94].

В [143] авторите обобщават характеристиките на потребителската грамотност в различните ѝ разновидности по следния начин:

- включване на нови умения, стратегии, наклонности и социални практики, които са задължителни в съвременния ИТ контекст;
- оптимизиране на възможностите на потребителите да използват дигиталните технологии;
- изключителна важност за участниците в глобалното интернет общество;
- постоянна промяна, свързана с бързото развитие на ИКТ.

Нарастващата потребност от компетенции за работа с (големи) данни през последните години налага и понятието **грамотност за работа с (големи) данни**. То се отнася до способността да се учим от данните. През годините, с напредъка в развитието на ИКТ, схващането за грамотност също се променя, както и технологиите за обработка и анализ на данни. Ако направим аналогия с общата грамотност, която включва способности за четене, писане и смятане, то при работата с данни може да се разграничат три вида дейности: *придобиване на данни* – четене; *представяне или споделяне на данни* – писане; *анализ и извличане* – генериране на нови знания, основаващи се на придобитата информация. Третата дейност може да се асоциира със способността за смятане в класическата концепция за грамотност, но тук работата с количествени данни не се ограничава до изследване на числови данни, а включва също прилагане на математическата логиката (дедукция и индукция) за извеждане на нови знания.

Грамотността за работа с големи данни обхваща и двете роли при работа с данни: пасивна и активна [55].

Пасивната роля се свързва с ученето от големите данни и получаването на полезна информация и знание от изследваните данни. Това включва използване на компютърни технологии, сложни аналитични техники и разнообразни техники за визуализиране и представяне в компактна форма на основните свойства на данните, както следва:

- Извличане на полезни данни в подходяща форма – **компетенции за извличане**. Включват се например разбиране на разнородния характер на събраните данни в организациите; използване на технологии за бази данни – SQL и NonSQL; прилагане на програмни езици, за да се уточни какво е необходимо и в какъв формат; изследване на съвременни технологии като хранилища за данни и разбиране на процеса „извличане – преобразуване – зареждане“ (*Extract-Transform-Load, ETL*), кубове с данни (*data cubes*) и системи за оперативна аналитична обработка (*OLAP*).

- Обобщаване и представяне на извлечените данни в полезна информация – **информационни компетенции**. В тази група се включват компетенции за разбиране и използване на технологии, включително инструменти за извличане и представяне на получените знания в подходящ формат. Това изисква разбиране за границите на възможностите на аналитичните техники. Освен задължителните знания и умения в области като статистика и математика е необходимо използването на различни форми за визуализация и способност за интерпретиране на извлечените данни.

- Генериране на полезни знания от придобитата информация – **компетенции за учене**. Прилагането на различни техники и описание

на методите, различни топологии за представяне на разстояние (distance) или подобие (likelihood) са начини да се обогати разбирането на това, което е скрито в данните, и да се даде възможност за улавяне на знанието. Умението да се разглежда критично получената информация в зависимост от съответствието ѝ с проблемната област, е ключово. Проверката на източниците и отчитането на ограниченията на прилаганите технологии, използването на подходящите аналитични инструменти за премахване или намаляване на рисковете, свързани с неразбиране или грешно разбиране, а също използването на техники за критична преценка на информацията са задължителни компоненти на тази компетентност.

- Изграждане на „мъдрост“ – **компетенции за развиване на опит.** Опитът се натрупва с течение на времето чрез експериментирание и извличане на нови полезни знания. Мъдростта представлява съчетание на жизнен опит, разум и интуиция и включва още проникателност, прозорливост, способност за вникване и знание за това кое е добро и правилно. Тя обединява знания и опит в прозрения и е свързана с разбирането на движещите сили на процесите и събитията и взаимовръзките между тях. Придобиването на мъдрост изисква компетенции като абстрактно мислене и способност за обобщаване, разграничаване на основните, екзистенциални фактори от вторичните, временни или специфични фактори.

Активната роля се свързва с процесите на предоставяне на данни на заинтересованите страни по ефективен и ефикасен начин. Може да различим няколко специфични компонента на грамотността при работа с големи данни:

- *Събиране на данни.* Днес технологиите позволяват събиране на огромни обеми данни в различни формати. В повечето случаи данните се записват в суров вид, без ясна представа как и дали ще бъдат използвани по-нататък. Във връзка с това възникват следните въпроси: „Какво може да се научи от данните?“, „Как може да се обработят?“, „Каква е тяхната стойност?“. Записаните данни представляват факти, събития, а също идеи, мнения и мисли, които се използват за извличане на полезното знание от тях с различна цел. Затова събирането на данни днес е важно умение.

- *Структуриране на данни.* Начинът на *организиране* на данните е пряко свързан с начините за тяхната обработка. *Аналитичните техники* могат да бъдат използвани за изследване и изучаване на данните. *Подготовката на данни* за по-нататъшното му използване е друго важно умение, което се изисква днес.

- *Съхранение на данни.* По отношение на съхранението на данните може да се разглеждат два основни аспекта: *местоположение* и (публичен,

частен) *достъп*. Местоположението обикновено е свързано със собствеността на данните, което поражда редица други проблеми. Облачните технологии предлагат нов, изключително удобен начин за съхранение на данни, но все още в тази инфраструктура има сериозни предизвикателства пред контрола на всички процеси по достъпа и сигурността на данните, което от своя страна рефлектира върху доверието на потребителите.

- *Достъп до данни*. Компетенциите, свързани със сигурността на данните и гарантирането на правилен достъп до тях, е друг компонент на грамотността за работа с данни. Във връзка с това възникват редица проблеми и предизвикателства, свързани със собствеността на данните, поверителността и други чувствителни зони по отношение на достъпа и преноса на данни.

Този кратък списък на специфични компетенции, необходими за работа с големи данни, ни позволява да говорим за „грамотност за работа с големи данни“. Повечето от тези компетенции са компоненти на общата цифрова грамотност, но в контекста на големите данни те носят нови измерения и изисквания към необходимите знания и умения. Днес получаването на тези специфични компетенции е предизвикателство за конкретния индивид, за социалните организации и за обществото като цяло, но най-вече за образователната индустрия.

Тези дейности изискват компетенции за използване на компютри за достъп и извличане и за записване и споделяне на данни, както и за анализиране на големи масиви от данни с помощта на сложни софтуерни приложения. По този начин грамотността за работа с големи данни се превръща в следващо ниво от еволюцията на „информационната грамотност“, изградена върху категориите „основна грамотност“, „цифрова грамотност“ и т.н. Грамотността за работа с големи данни определя и ново измерение на понятието „грамотност“, което се отнася до познавателния капацитет на човека и способността му да се учи от данни с количество и сложност, които са значително над човешките природни възможности.

Ученето от големи данни се сблъсква със значителни трудности. Основната трудност идва от неспособността да се наблюдава директно целият набор от свойства на обектите поради техния обем. Обектите може да се наблюдават само чрез обобщаващи статистики. Валидността на получената информация зависи от това дали данните могат да удовлетворят определен набор от изисквания, например дали различните параметри са взаимно независими. Доказването на независимост обикновено е трудна задача, а приемането на независимост без доказателства може да доведе до заблуждаващи и погрешни изводи.

Трябва да бъдат решени три категории задачи:

(1) да се открият, извлекат и използват данни, отговарящи на определени изисквания, с оглед получаването на валидни резултати при прилагането на дадена аналитична техника;

(2) да се провери дали данните отговарят на тези изисквания;

(3) да се изследва въздействието на неудовлетворените или частично удовлетворени изисквания върху получените резултати и това разбиране да се отнесе към решението на проблема.

1.3.2. Дигитално неравенство

Дигиталното неравенство е термин, който се отнася до разликата между лицата, изграждащи дадено общество както в демографски, така и в регионален аспект. Най-често се свързва с липсата на достъп до съвременни ИКТ, включително телефон, телевизия, персонален компютър и интернет.

Въпреки че при появата си терминът се свързва с неравенството между „имащите“ и „нямащите“ интернет достъп, в съвременното общество все по-често се наблюдава по-сложна социална, икономическа и културна диференциация. Редица фактори все още влияят върху индивидите и възможностите им да използват съвременните технологии.

Днес дигиталното неравенство в обществото се поражда от ускореното развитие на ИКТ и постоянната поява на нови техники, методи и технологии с данни. Може да се разгледат различни негови аспекти. Организациите имат възможност да оползотворят данните, генерирани от хора и машини, с цел подобряване на ефективността и ефикасността на операциите [136]. В много случаи хранилищата за данни са проектирани по такъв начин, че да не позволяват лесен достъп и изследване на натрупаните данни. Това, наред с липсата на грамотност за работа с данни, са фактори за възникването на нова форма на дигитално неравенство, свързана с използването на големи данни и извличането на полза от тях (*Big data divide*). Разделението в обществото обаче не е просто онова, което Boyd и Crawford (2011) описват като базирано на „имащи“ и „нямащи“ достъп до интернет [45]. Полученото неравенство по отношение на използването на генерираните от хора и машини големи данни поражда въпроси, свързани и с асиметрични процеси на сортиране и различни начини на мислене за това как данните да се свържат със знанието и неговото приложение [27, 28] в контекста на конкретния проблем, който изисква решение.

След проучване на научната литература, свързана с този проблем, в следващата Таблица 1.8 е направен опит да бъдат представени съществуващите дефиниции и гледни точки на изследователите.

Таблица 1.8. Дигитално неравенство в контекста на големите данни

Автор	Становище
Lyon, 2002	Лицата, които притежават достъп до данни, експертиза и способност за обработка, имат възможност да участват във все по-сложни и непрозрачни форми на сортиране, които могат да бъдат „мощно“ средство за създаване и засилване на дългосрочни (или новосъздадени) социални различия [149].
Weinberger, 2011	Малцина ще имат достъп до полезни форми на „знание“, осигуряващо предимство. Тези форми ще останат не само недостъпни, но и неразбираеми за огромното мнозинство [236].
Boyd and Crawford, 2011	Разделението по отношение на големите данни съществува между Big Data rich (включително компании и университети, които могат да генерират или купуват и съхраняват големи набори от данни) и Big Data poor (лица, нямали достъп до данни, експертиза и способност за обработка) [44].
Andrejevic, 2014	Разделението в областта на големите данни представлява асиметричната връзка между тези, които събират, съхраняват и обработват големи количества данни, и тези, към които е насочено събирането на данни. Това е разделение между онези, които са в състояние да извлекат и използват непредвидими и необясними констатации, и лицата, чието съществуване е за сегнато от произтичащите от това решения. Големите данни ще бъдат достъпни само за тези, които имат ресурси за поддържане на най-новите технологии и най-големите бази данни [27, 28].
Espinosa et al., 2015	Днес сме свидетели на нов сценарий на дигитално неравенство, при който неспециалистите не са в състояние да се възползват от големите данни. Успешното извличане на данни изисква ноу-хау на експерт, за да се придобият надеждни и полезни знания в получените модели [95].
Christozov and Toleva, 2015	Големите данни установяват ново разделение на обществото между онези, които са способни да се учат сами от тях (членове на т.нар. Big Data Elite) и тези, които разчитат на посредници, за да „изучават“ данните. Това ново разделение добавя нови аспекти на съществуващото в обществото дигитално неравенство. За да се получи знание от натрупаните сложни данни, са необходими подходящи методи и технологии. Това от своя страна ограничава броя на хората, които притежават необходимия опит, за да се възползват от източниците на данни [55].

Все по-важно става разбирането на наличните възможности за съхраняване, обработка и търсене в големи масиви с данни, но още по-важно е наличието на умения за извличане на полезните знания от

данните и тяхното използване. Критичното изучаване на данните, трансфера на знания и изграждането на дълбоко разбиране на процесите са от съществено значение за формирането на рационално поведение. Запазването на знанието е друг съществен аспект, който оказва влияние върху рационалното поведение.

В повечето дефиниции за големи данни се споменава, че наличието на повече данни предполага повече възможности за анализ, а на свой ред те са основа за вземането на по-информирани решения. Казаното важи с пълна сила само за експертите в анализа на данни, както и за онези компании, които могат да си позволят да ги наемат. За останалата част от обществото тези ползи ще останат недостъпни.

1.3.3. Измерения и измерители

Измеренията на дигиталното неравенство в контекста на големите данни, проявяващи се във и извън организациите, може да се разглеждат на няколко равнища, както следва:

- **Фирмено/организационно равнище** – разделение между фирми и организации, използващи натрупаните данни и извличащи конкурентно предимство спрямо фирмите и организациите, които не го правят. Форма на разделение може да се наблюдава и между фирмите, които разполагат с човешки капитал с по-добри аналитични умения, и онези, които не разполагат с такъв.

- **Вътрешнофирмено равнище** – разделение между ИТ специалистите, които могат да се учат от големите данни, и останалите, които само могат да ги управляват и модифицират.

- **Социално равнище** – разделение между лицата, които притежават умения и могат да боравят с големи данни, извличайки полезна информация от тях, и онези, които се нуждаят от медиатор, за да се възползват от големите данни.

Освен това разделение на обществото по отношение на работата с големите данни може да се наблюдава и на национално, и на глобално равнище.

Въз основа на извършения преглед на литературата ние смятаме, че неравенството при работа с големи данни се наблюдава като резултат от следните фактори, които ще бъдат разгледани по-долу:

А. Липса на достъп до данните, както и липса на технологии и инфраструктура достъпните данни да бъдат използвани.

Б. Липса на умения и възможност за извличане на знания/прозрения от натрупаните данни (грамотност за работа с големи данни).

В. Липса на специалисти с нужния опит и експертиза за работа с големи данни.

Г. Липса на информираност относно потенциалните ползи и възможности на обработката на големи данни.

Всеки от горепосочените фактори ще бъде разгледан поотделно.

А. Достъп до данни и инфраструктура

Основните фактори за формиране на дигитално неравенство по отношение на използването на големи данни според Andrejević са финансовите измерения [27, 28]:

- използването на големи данни изисква достъп до и контрол върху скъпа технологична инфраструктура;
- поддръжката на набор от данни, както и софтуер, процесорна мощ и експертен опит за анализирането им имат висока, а с разрастването на пазара – и все по-висока цена.

Базирайки се на сравнението, направено от Weinberger, че в ерата на големите данни „най-умният човек в стаята е самата стая“, Andrejević внася опасения, свързани с това кой притежава и управлява „стаята“ [236, 27, 28]. На практика „знанието“, получено чрез интелигентен анализ на данни, е достъпно само за тези, които имат достъп до машините, базите данни и алгоритмите.

Ерата на големите данни се характеризира с възможността да се използват особено големи бази данни и предлага нови възможности за вземане на решения и правене на прогнози, които обаче остават недостъпни за лицата, нямащи достъп до базите данни, хранилищата и мощността на компютрите. Това поставя в по-изгодна позиция онези, които имат достъп до споменатите по-горе ресурси за извличане на данни, в сравнение с тези, които нямат такъв достъп.

Б. Грамотност за работа с големи данни

Явлението „големи данни“ поставя нова линия на разделение между хората по фактори като грамотност и компетентност [54].

За работа с големи данни е необходимо да се притежават способности за:

- търсене, идентифициране и извличане на данни, свързани с даден проблем;
- използване на различни техники за проверка на надеждността и релевантността на получените данни;
- използване на различни техники за представяне на огромно количество данни смислено и сравнимо с когнитивния капацитет на

потребителите, способност за разбиране на специфични ограничения, изисквания за приложимост и качество на информацията, и интерпретиране на информацията, генерирана чрез тези техники.

Тези компетенции помагат да се разберат данните, представляващи имплицитни свойства на обекти или събития, и да се подобри вземането на решения [54].

Според Mikalef et al. уменията за работа с данни са може би най-търсеният ресурс в компаниите, които разполагат с големи данни, тъй като специалистите по данни подпомагат задаването на правилните въпроси и преобразуването на данните в практически прозрения [162, 163]. Очевидно софтуерът, инфраструктурата и данните са недостатъчни, за да осигурят каквато и да е стойност, ако не са налични експерти с подходящи персонални умения и знания, за да ги приложат в действие.

В. Специалисти, притежаващи нужната експертиза да се учат от данни

Рамката на знанията и уменията, които трябва да притежава съвременният специалист по данни, продължава да се разширява. Наблюдава се постоянен растеж и в потребността на бизнеса от квалифицирани специалисти, притежаващи нужната експертиза.

Според Christozov&Toleva големите данни ни изправят пред ново предизвикателство, което е насочено към способността на хората да откриват знания в количество данни, значително надхвърлящо човешките когнитивни възможности [55]. Може да се твърди, че е трудно да се постигнат необходимите компетенции за придобиване на грамотност за работа с големи данни. Тенденциите в образователно развитите страни показват, че младото поколение се отдръпва от изучаването на теми, свързани с анализа на данни, като математика и статистика. Тези поколения разчитат на медиатори – хора (информационни брокери) [55] или ИТ, обикновено без необходимото разбиране за ограниченията при прилагането на инструментите и нивото на релевантност на резултатите спрямо същината на проблема. Този начин на изследване на големи данни не генерира подходящи знания за обектите и събитията, описани от данните. Само определен елит ще може напълно да се възползва от натрупаните данни, да разбере причинно-следствените взаимоотношения в процесите, което ще му позволи предсказване на тенденциите [29] и влияние върху поведението на системите и обществото.

Според [36] уменията на ИТ персонала са критичен ресурс за натрупването на измерима бизнес стойност. Представата, че фирмите

трябва да обединяват технологията с уменията, притежавани от човека, стои в основата на социално-техническата рамка на науката за данните. Това повишава ролята на човешкия фактор и показва, че максималното технологично представяне изисква едновременно управление и подхранване на човешките умения и знания на организациите [233]. Тази идея е особено важна в контекста на големите данни, тъй като уменията не са нужни единствено при експлоатацията на технически ресурси (например софтуер и инфраструктура). По-важно е генерирането на прозрения, които ръководят вземането на решения в организациите [202].

Г. Информираност относно потенциалните ползи от големите данни

Много често организациите не са достатъчно добре запознати с технологичните промени и новите възможности, които те предоставят за повишаване на конкурентоспособността им [146]. На практика това поставя фирми, които не се „учат“ от данни, в неравностойно положение спрямо други, които трупат „бизнес стойност“, базирана на данни. Без подходящи инструменти за извличане на полезното знание, скрито в големите данни, организациите ще продължават да са в позиция на „догонвачи“ своите конкуренти на пазара.

Освен четирите основни фактора, които бяха разгледани по-горе, съществуват и други, второстепенни фактори, влияещи върху разделението на обществото в контекста на големите данни. Например според Sukier извличането на данни води до загуба на „степената на прозрачност поради обстоятелството, че компютрите сами вземат решенията“ [69]. Според автора съществуват машини, които се обучават. Това позволява на машините да калибрират връзките между различни аспекти на информацията, което няма как да бъде забелязано от човешкия фактор [27, 28]. Машинното обучение има две страни. От една страна, машините подпомагат хората в разбирането на процесите и явленията, като позволяват извлечените от данните зависимости да бъдат представени в разбираема за хората форма. Такъв е случаят със статистическите методи за машинно обучение (*Statistical Machine Learning*). От друга страна, машините използват алгоритми, симулиращи поведението на човешкия мозък, като например изкуствените невронни мрежи, и позволяващи на компютъра да следва интелигентно поведение, да взема обосновани решения и да решава проблеми, да се дава експлицитна форма, разбираема за хората. Развитието на тази страна на алгоритмите и методите за обработка на данните е свързано с термина *Deep Learning*.

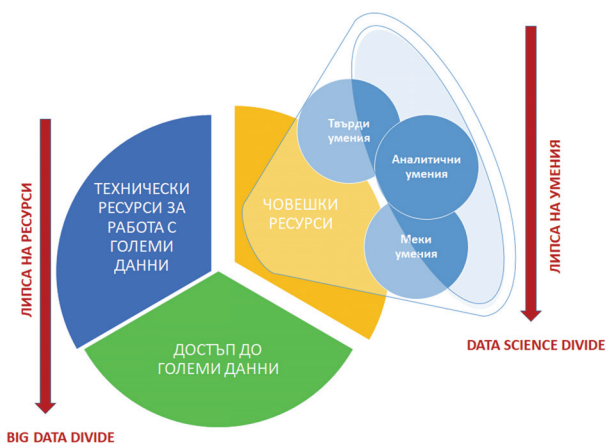
1.3.4. Неравенство по отношение на науката за данните

По своята същност разделението в сферата на големите данни може да бъде представено като разделение, породено от недостига или липсата на един или няколко от необходимите ресурси за работа с големи данни. Ние дефинирахме следната рамка от три основни компонента (Фиг. 1.6):

- (1) технически ресурси;
- (2) достъп до големи данни;
- (3) човешки ресурси.

Наличието на всеки от тези компоненти при създадени условия за тяхното нормално функциониране гарантира конкурентоспособност на организацияте, а отсъствието или незадоволителното функциониране на някои от тях би създавало условия за формиране на неравенство.

Интересен за нашето изследване се оказва компонентът „човешки ресурси“, изискванията към който поставят основите за формиране на подразделение на съществуващото неравенство при използването на големи данни. Човешкият фактор като ресурс (специалист по данни) следва да притежава различни категории умения. Липсата на някой от тези компоненти, както и недостигът в умения за работа с големи данни биха създали условия за формиране на ново разделение, което ще наричаме *дигитално разделение в областта на науката за данните (Data Science Divide)*.



Фиг. 1.6. Дигитално разделение в сферата на големите данни и науката за данните

С други думи, това разделение в контекста на науката за данните показва различия на ниво човешки капитал и разделение на ниво умения за работа с големи данни, създаване на хипотези, осъществяване на тестове, анализ на резултатите, съпоставка с очакваните стойности, както и вземане на решения на базата на получените резултати.

Ние ограничаваме списъка с предизвикателства пред науката за данните до списък с четири основни:

- *Данни.* Първото предизвикателство при работата с големи данни е свързано с характеристиките на събраните масиви от данни и извличането на полезна информация с добро качество, точност, достоверност и др.

- *Технически средства и алгоритми за анализ.* Обработката, съхранението и трансферът на големи масиви от данни създават предпоставки за възникване на чисто технологично предизвикателство, явяващо се второ по ред в нашия списък с предизвикателства.

- *Компетентност и експертиза на анализаторите.* Специалистите, анализиращи големи данни, се стремят да извлекат скрити смислени модели, вътрешни връзки и зависимости, които не може да се установят непосредствено от натрупаните данни. Те са изправени пред много нови възможности и предизвикателства, които им предоставят големите данни, наличните методи и технологии за тяхното съхраняване, обработка и анализ.

- *Сигурност на данните.* Често потребителите на услуги и устройства, генериращи данни, не са информирани, че ще бъдат включени в голямото количество данни и/или за какво тези данни може да се използват впоследствие. Четвъртото предизвикателство се отнася до някои етични аспекти при анализа на големи данни, които засягат поверителността и правните проблеми, свързани с авторското право и собствеността на данните.

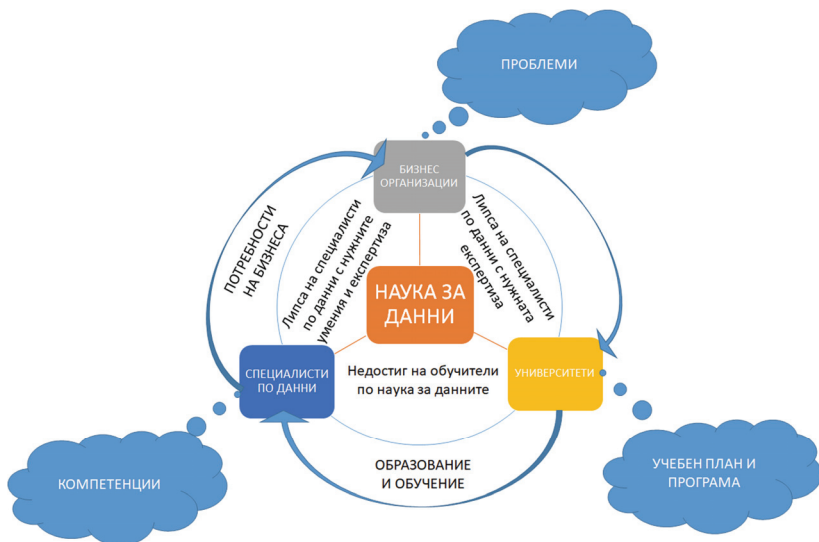
Във връзка със създаването и реализацията на специалисти, подготвени да работят с големи данни, ние отграничихме три допълнителни предизвикателства в областта на науката за данните:

- *Предизвикателства пред обучителните организации, изграждащи специалисти за работа с големи данни* – университетите трябва да отговарят на моментните потребности на пазара на труда и да предлагат адекватни учебни програми, които да дават възможност на завършващите специалисти да се реализират по професията си веднага след своето дипломиране. Това от своя страна налага ангажирането на преподаватели с опит и експертиза, съответстващи на пазарното търсене. Естествените въпроси, които възникват, са: „Какви знания, умения и

компетенции са необходими за извличане на полезна стойност от натрупаните данни?“ и „Каква е честотата, която трябва да бъде следвана при одит и актуализация на съществуващия учебен план?“.

- *Предизвикателства пред бизнес организациите и институциите, които са собственици на големи данни* – предизвикателствата тук се свързват, от една страна, с откриването и назначаването на експерт по големи данни, притежаващ знанията и опита да се справи с наличните такива данни, а от друга страна, с процеса на съхраняването на натрупания актив от големи данни. Второто предизвикателство налага грижа за целостта и сигурността на данните и допълнителни ресурси (технически и финансови), които да позволят съхраняването и последващия анализ на данните.

- *Предизвикателства пред специалистите в областта на науката за данните* – работата с големи данни изисква постоянно надграждане и усъвършенстване на знанията, уменията и компетентностите. Желанието за развитие в тази посока е задължително условие за поддържане на експертното ниво на специалистите в областта на големите данни.



Фиг. 1.7. Цикъл на развитие на науката за данните

Динамиката на развитие на науката за данните зависи от комуникацията между три страни (Фиг. 1.7): университетите, обучаващи специалисти, готови да анализират големи данни; бизнес организациите,

търсещи способни специалисти, и самите специалисти по данни. Липсата на стабилна връзка между тези три страни ще доведе до диспропорции, които ще се отразят върху темпа на развитие на науката за данните. На практика университетите имат отговорността да подготвят специалисти, отговарящи на търсенето от страна на бизнеса. За да се осигури висока стойност на процеса на обучение, в подготовката на студенти трябва да бъдат привлечени преподаватели с нужната експертиза. Липсата на подобни експерти ще се отрази върху качеството на учебния процес. Това би довело до недостиг на способни специалисти по данни и до производство на кадри с недостатъчен набор от знания, умения и компетентности, за да се реализират успешно на пазара на труда. Липсата на експерти в областта на големите данни ще отвори нова форма на дигитално неравенство на фирмено равнище, разделящо фирми, които имат човешки ресурс, способен да извлича знания от данни, и фирми, които нямат нужните специалисти за това.

На практика осигуряването на кадри е задача на университетите и обучаващите организации. Основният проблем тук е свързан с възможностите за адаптация и поддържане на равен темп на предлагането с този на търсенето. Огромно предизвикателство е възможността за своевременно актуализиране на съществуващите учебни планове в съответствие с промените на пазара на труда.

1.3.5. Подходи за преодоляване на неравенството по отношение на големите данни

Изучаването на големите данни може да отговори на ключови предизвикателства и да подпомогне вземането на по-ефективни решения, базирани на прозрения и доказателства. Освен това предоставя възможност за конвертиране на сложни, често неструктурирани данни в приложима информация, като стратегически отговор на променящите се глобални тенденции [70].

С цел търсене на начини за справяне със съществуващото разделение в ерата на големите данни, в следващите подточки ще бъдат разгледани някои ключови фактори.

А. Осигуряване на достъп до технологична инфраструктура

Според Longbottom ръководството на всяка организация трябва да се увери, че разполага с подходящата техническа платформа, за да се справи с настоящите и бъдещите си информационни потребности

[146]. За тази цел е необходимо да се осъществи преглед, който дава възможност да се определи кои от съществуващите елементи в наличната инфраструктура може да се използват и какви нови елементи трябва да се въведат, за да бъде осигурена рентабилна платформа за управление на данните (в частност, знанията за бъдещето).

Б. Осигуряване на user-friendly технически средства

Извличането на данни е сложен процес, съставен от набор от стъпки, които трябва да бъдат приложени към източниците на данни, за да се открият знанията в тях. Една от причините, които възпрепятстват прилагането на техники за извличане на данни, е, че потребителите, които не са експерти в науката за данните, не са в състояние да изразят своите изисквания за извличане на данни, т.е. какъв вид знания може да се открият в данните [95].

Когато става въпрос за задоволяване на съществуващи технологични потребности, на първо място в списъка с приоритети трябва да бъде лекотата на употреба. Инструмент, който е насочен само към хора с висока грамотност, може да се използва само от тях. Според Longbottom това би могло да се разглежда като опасност, тъй като знанията и уменията на специалистите по данни им предоставят уникални възможности, които другите потребители не притежават [146]. Следователно възниква нужда от създаване на подходящи технически средства и инструменти, които да дават възможност на потребителите да обработват и анализират данни по удобен и лесен за тях начин. За тази цел е необходимо интерфейсите на използваните продукти да са интуитивни и удобни за работа. Освен това визуализацията на данните трябва да бъде ефективна и гъвкава, така че да отговаря на разнообразните изисквания и предпочитания на потребителите.

Според Espinosa et al. е необходима демократизация на *Data Mining*, осигуряваща подходящи техники за извличане на данни и настройки според характеристиките на данните [95]. В подкрепа на казаното Kriegel твърди, че удобния за потребителите начин за извличане на данни (*user-friendly Data Mining*) е крачка напред към необходимата демократизация, тъй като насърчава откриването на знания без овладяване на концепции и техники за извличане на данни, преодолявайки съществуващото разделение в контекста на големите данни, и позволява на всички да се възползват от наличните големи данни [137].

В свое изследване Espinosa et al. дават предложение за включване на неспециалисти в областта на науката за данните в процеса на проектиране на системите за извличане на данни [95]. Това са потребители, които нямат задълбочени познания, свързани с извличането

на данни, но имат визия за начина, по който да се осъществяват основни процеси и дейности, имат свои изисквания и препоръки. Предложеният от авторите подход включва набор от механизми за насочване на неспециалистите към интерпретиране и използване на резултатите от процеса на извличане на данни.

В. Осигуряване на подходящо обучение на специалисти в областта на науката за данните

Ефективната обработка и анализ на големите данни изискват специалистът да притежава задълбочени познания в областта на статистиката, умения за прилагане на статистически методи чрез използване на сложен софтуер и обширни познания в областта на науката за данните, както и в областта на компютърните науки и технологии. Тези компетенции имат решаващо значение във всяка област, а съчетаването на изброените аспекти в обучението създава значителни предизвикателства пред университетите и обучаващите организации [54].

РЕЗУЛТАТИ И ИЗВОДИ ПО ПЪРВА ГЛАВА

Глава I представя основополагащите понятия в сферата на науката за данните, като „големи данни“, „грамотност за работа с (големи) данни“, „дигитално неравенство“ и др.

Извършеното комплексно изследване по отношение на проблемите и предизвикателствата в ерата на големите данни ни дава основание да формулираме следните резултати и изводи по Глава I:

1. Направен е преглед на литературата, отнасяща се към дефиниране на кръга от проблеми, свързани с науката за данните.

2. Изведени са основните характеристики на проблемите, които големите данни пораждат, като социални, методологически и приложни.

3. Направено е обобщение на получените резултати, като са изведени структурни модели на обхвата и насоки на предизвикателствата и проблемите, които се очаква да решават анализаторите на данни – специалисти по науката за данните.

ГЛАВА II. ПРЕГЛЕД НА ЗАДЪЛЖИТЕЛНИТЕ КОМПЕТЕНЦИИ В ЕРАТА НА ГОЛЕМИТЕ ДАННИ

В ерата на информацията „наука за данните“, „големи данни“ и „информационно общество“ са често използвани термини, които отразяват важни аспекти, свързани с голямото количество данни и информация, до които днес имаме достъп. Триадата „данни – информация – знание“ характеризира процеса на информационно осигуряване и учене от данните с оглед на ефективното им използване.

Днес организациите разполагат с огромни масиви структурирани и неструктурирани данни, отразяващи изцяло дейността им, от които може да се извлече полезно знание за повишаването на тяхната конкурентоспособност в условията на глобален и силно конкурентен пазар. За да се извлече полезно знание от големи данни, са необходими технологии, софтуер и компетентни професионалисти – анализатори. Във връзка с предизвикателството „големи данни“ и необходимостта от специалисти, които да подпомагат процесите на вземане на решения, възниква нова професия – специалист по данни (Data Scientist).

Естественият въпрос, пред който са изправени образователните институции, е какви знания, умения и компетенции са необходими за извличане на полезна стойност от натрупаните данни. За да дефинираме професионалния профил на съвременния специалист по данни, в тази глава представяме и анализираме резултатите от проучване и анализ на релевантните литературни източници (1) и на търсените умения от страна на бизнес организациите (2).

2.1. КОМПЕТЕНТНОСТЕН ПОДХОД

В образователната система съществува схващането, че има основно ядро от умения и познания, които всички професионалисти от една област трябва да притежават, за да изпълняват ефективно съответните дейности.

При управлението на човешките ресурси се използва подход, който се основава на идентифициране, дефиниране и измерване на индивидуалните различия по отношение на способностите за успешно изпълнение на дадена работа. В редица международни компании

приложението на такъв **компетентностен подход** при управлението на хора се превръща в интегрирана политика, обхващаща повечето функции – от планирането на човешките ресурси, набирането и подбора им, през обучението и развитието, управлението на изпълнението на работата, системите за възнаграждения и мотивация, до промените в организационната култура. Днес компетентностният подход се възприема, разработва и усъвършенства от международната научна общност като перспективен път за повишаване на качеството на образованието и на живота на съвременния човек [3].

В литературата съществуват различни дефиниции на „компетентност“:

- вътрешни характеристики, които са решаващи за ефективно изпълнение на дадена задача, роля или ситуация [211];
- поведенчески аспекти, които влияят на ефективността на работата;
- свойства, знания, навици и ценности на личността, които са свързани с ефективното изпълнение на дадена работа.

„Компетенция“ също е понятие, използвано широко в областта на обучението, развитието и управлението на хора. Често понятията „компетенция“ и „компетентност“ се смесват или използват като синоними, но всъщност те имат различен обхват [1, 2]. Компетентността (*competency*) представлява интегрално множество от знания, умения и отношения (подходи), използвани за решаването на даден проблем в определен контекст. То описва аспекти на възприятието, разбирането, целепологането и поведението, които стоят зад обмислянето на изпълнението на работата (на базата на възгледите, светогледа и свързаните с тях професионални знания, умения, навици). Компетенцията (*competence*) е понятие, свързано с обхвата на професионалната дейност, която едно лице е компетентно да извършва, и отразява неговия статут в организацията, правомощията и областта на отговорност. Във връзка с това компетентността в дадена област се разглежда като съвкупност от компетенции.

По своя характер компетентността е интегративна, има пряко отношение към качествата и ценностната система на личността и е с практико-приложна ориентация. Именно в това се свежда и основната разлика между компетентността и понятията „знания“ и „умения“. Те имат общи свойства и елементи, но и съществени различия (Фиг. 2.1).



Фиг. 2.1. Триада „знания – умения и навици – компетенции“
(адаптирана от [8])

Научните знания включват факти, понятия, закони, закономерности и теории. Придобиването на знание е когнитивен процес, ориентиран към събиране, разбиране, осъзнаване и приемане на информацията на ниво семантика и прагматика и е резултат от обработената информация. Типични въпроси за данните и информацията са „Кой?“, „Какво?“, „Къде?“ и „Кога?“, докато въпросите за знанията са „Как?“ и „Защо?“. Знанието не е константна величина, то се променя непрекъснато в процеса на учене в ежедневието и на работното място. Освен това то е ситуационно и изисква не само процедура и механизъм, но и подходящ момент и обстоятелства, за да бъде приложено.

От когнитивно-психологическа гледна точка може да бъдат разграничени различни видове знания, като най-широко се използва разграничението между **декларативно** и **процедурно** знание. Декларативните знания са свързани с факти, събития, понятия, дефиниции и т.н. от дадена област, които човек знае и може да споделя и разпространява. Процедурните знания включват знания за осъществяването на различни дейности и начините за решаване на задачи в дадена област (know-how). Те са свързани с използването на декларативни знания и обикновено се отнасят до знания, които не може да бъдат съобщени.

Знанието условно се разделя на **явно (експлицитно)** и **неявно (имплицитно)**. Явното знание е осмислено знание, изразено в мате-

риална, най-често словесна форма и може да се предава във формиран вид на различни носители (документи, инструкции, книги и др.). То може да бъде систематизирано, описано и разказано на други хора. Неявното, скрито знание е знанието, притежавано от отделния индивид и определящо неговото поведение. Неявното знание често не може да бъде предадено във вербална или друга експлицитна форма. Понятието „скрито знание“ често се използва в контекста на експертното знание на професионалистите в дадена област.

Умението е усвоен в резултат на повторение или по подражание начин за самостоятелно ползване на знания и понятия, умствени процеси и физически действия за решаване на теоретични и практически задачи. Умението е придобито в процеса на труда и общуването, познанието и самопознанието възможност за действие или поведение в дадена ситуация, която предполага използване на подходящ алгоритъм за постигане на желана цел (резултат) и в този смисъл си прилича със способностите [8]. Уменията и способностите се отнасят към по-общото понятие „възможност за решаване на задачи“. В структурата на способностите се включват уменията, тъй като, ако определено умение не се утвърди като трайно притежание на личността, то става компонент на нейните способности.

Навикът и умението имат общи белези – изграждат се в резултат на многократни повторения и упражнения и съставят изпълнителската част на всяка дейност. Навикът представлява затвърден начин за изпълнение на действието с предмет или инструмент в резултат на многократни повторения и (или) упражнения. От психологическа гледна точка навиките се делят на: *сензорно-перцептивни* (за възприемане, моторни или двигателни и интелектуални), и *умствени* (за решаване на задачи).

В Таблица 2.1 са посочени някои основни различия между гореизложените понятия [19].

Таблица 2.1. Сравнение на основните понятия

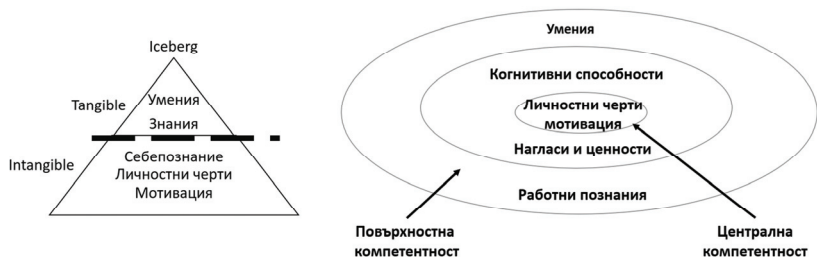
Способности, знания, качества, отличителни черти	Компетенции
<ul style="list-style-type: none"> • Базисни • Характерни за новонаети служители • Присъщи на входящото ниво на работа • Обикновено, за което плащат организациите. 	<ul style="list-style-type: none"> • Изискващи мултиплициране на знания, способности и персонални (личностни) черти. Придобиват се от служители с богат трудов опит (експерти). • Проявяват се в процеса на работа. • Необходимо е навлизане в дълбочина в работата. • Това, за което би трябвало да плащат организациите.

Завършващият висше учебно заведение (ВУЗ) трябва да притежава определени компетенции, наречени **ключови** – професионално релевантни качества и потенциал, които ще бъдат актуализирани в процеса на осъществяване на професионалната дейност и ще свидетелстват за неговата компетентност. Ключовите (базови) компетенции са [8]:

- политически и социални компетенции (способност за лична отговорност, участие в съвместно взети решения, регулиране на конфликти по мирен път);
- компетенции, свързани с живота в многокултурното общество (приемане на различията, уважение към другите, способност за взаимодействие с хора от други култури, езици, религии);
- компетенции, отнасящи се към овладяването на устно и писмено общуване на повече от един език;
- компетенции, свързани с възникването на информационното общество (овладяване на нови технологии, разбиране на възможностите на техните приложения, критично отношение към информацията, разпространяваща се със средствата на масовата информация и реклами);
- компетенции за нововъведения;
- компетенции за учене през целия живот на базата на непрекъснатото обучение, в контекста на професионалната подготовка.

Ключовите (базови) компетентности са многофункционални, надпредметни, междудисциплинарни, основават се на определено ниво на интелектуално развитие и са многомерни, т.е. включват различни личностни качества, интелектуални способности и комуникативни умения.

Spencer and Spencer разглеждат формирането на компетенциите във вид на айсберг, където знанията и навиците, които сравнително лесно може да се усвоят, съставят неговата видима част, а личностните черти, мотиви и Аз-концепцията са скрити „под морското равнище“, тъй като те трудно се развиват в процеса на обучение (Фиг. 2.2) [211].



Фиг. 2.2. Модел на компетентност „айсберг“ (Spencer & Spencer, 1993)

Според S. Parry, 1996 компетенциите трябва да се разделят на „твърди“ и „меки“ [178]. Твърдите компетенции се отнасят към професионално-специфичните особености (знания и умения), а меките компетенции – към личностните черти, ценности и стилове.

Компетентността може да се разглежда на две нива в зависимост от обхвата ѝ: **индивидуална** и **организационна**. Според Wagner организационната компетентност е не просто съвкупността от компетентностите на отделните индивиди, а сложна комбинация от индивидуалните компетентности, върху които оказват влияние редица стратегически, структурни и културни фактори [230].

За целите на нашето изследване ще разгледаме индивидуалната професионална компетентност и включените в нея ключови компетенции на специалистите по данни.

2.2. ПРОФЕСИОНАЛЕН ПРОФИЛ НА СПЕЦИАЛИСТА ПО ДАННИ

2.2.1. Специалист по данни

През последните години науката за данните се утвърди като мулти-, интер- и трансдисциплинарна област, в която се включват много и разнообразни области. Специалистите по данни се занимават с анализ на данни и извличане на скрити смислени модели, вътрешни връзки и зависимости, които не може да се получат непосредствено от натрупаните данни. Те са изправени пред много нови възможности и предизвикателства, които им предоставят големите данни, наличните методи и технологии за тяхното съхраняване, обработка и анализ, и не на последно място, оценяването на получените резултати в контекста на конкретната проблемна област. Естественят въпрос, който възниква, е какви знания, умения и компетенции са необходими за извличане на полезна стойност от натрупаните данни.

С оглед на комплексния характер на науката за данните специалистите по данни трябва да притежават комбинация от умения и компетенции в областта на програмирането, анализа, статистиката, моделирането и комуникацията. Те използват различни източници на разнообразни данни, чиито характеристики са от съществено значение в процесите на тяхното изследване и анализ, с цел извличането на полезна информация с добро качество, точност, достоверност и др. Затова те трябва да притежават обща компетентност за работа с данни и да отговорят на следните въпроси:

- Какво е качеството на данните? Как да се провери? Как да се осигури? Какъв е ефектът от „частично“ удовлетворяване на критериите за качество?

- Какви са съответните критерии за оценка на качеството на данните, получени от даден източник на данни и отнасящи се към съответната проблемна област?

- Какви са факторите и обстоятелствата, които влияят върху събирането и представянето на данни по определен начин?

- Как ще се използват данните?

- Как да се измери дали наличните и извлечените данни отговарят на горните критерии, за да се гарантират значими изводи?

Събирането на данни в реално време поставя нови възможности и предизвикателства за разбирането на данните, например комбиниране на административни данни с големи данни, идващи от различни източници, като търговски данни, мобилни устройства и сензори, социални медии и други обществено достъпни данни. Това, от една страна, изисква нови средства и методи, а от друга – нови умения и визия за работа с данни.

Специалистите по данни използват съвременните методи, техники и технологии за работа с големи данни, описани в Глава I. Задачите на специалистите по данни може да се групират в следните пет категории [224]:

- дескриптивен анализ (описание: Какво се случи?);

- диагностичен анализ (преценка: Защо се случи?);

- прогностичен анализ (прогноза: Какво ще се случи?);

- предписващ анализ (препоръка: Какво може да се направи/промени?);

- превантивен анализ (препоръка: Какво още може да се направи?).

Дескриптивният анализ (*Descriptive Analytics*) е свързан основно с установяването на никакви присъщи характеристики и връзки, определящи състоянието на наличните до момента данни и поведението на процесите, които те описват. Техниките за описателен анализ категоризират, характеризират, консолидират и класифицират данни, за да ги преобразуват в полезна информация за разбиране и анализ на текущи и минали резултати. Използват се статистически анализ на данните, клъстеризация, асоциации и др. Обобщените данни може да бъдат визуализирани в смислени диаграми и отчети, например за бюджети, продажби, приходи или разходи [131].

Диагностичният анализ (*Diagnostic Analytics, Inquisitive Analytics*) е

проучвателен анализ на съществуващите или допълнителни данни, който има за цел изследване и оценяване на влиянието на различни фактори и причини върху дадено следствие. Освен това се използва за разкриване на несъответствия, диспропорции, противоречия и проблеми, опит за контрол на някои смущаващи фактори. Примери за такъв вид анализ са статистическият анализ, факторният анализ и др. [38].

Прогностичният анализ (*Predictive Analytics*) се отнася до обосноваване и конструиране на различни форми на научни предвиждания, разкриване на перспективи, но също и на възможни опасности и проблеми в бъдещото изменение на данните и параметрите на процесите въз основа на наблюдаваните им стойности и динамика до момента. Предвижданията се извършват под формата на: (1) предположения (допускания) какво ще се случи с даден обект в близкото или по-далечното бъдеще; (2) различни варианти на прогнози; (3) разработване на план/програма. Използват се задачите за класификация, регресионните статистически модели, анализ на динамични редове (*time series*) и машинно обучение за намиране на смисъл в големи количества данни [231].

Предписващият анализ (*Prescriptive Analytics*) представлява изследване на различните алтернативи във връзка с прогнозираните резултати с цел подпомагане на анализаторите да оптимизират своя подход за постигане на най-добър отговор или действия, необходими за постигане на бизнес целите. Извършват се изследвания и симулации, за да се проследят последиците на различните сценарии (*what-if анализ*) и да се издаде препоръка за действие [131].

Превантивният анализ (*Pre-emptive Analytics*) е свързан с възможността за предприемане на предпазни мерки за събития, които могат да имат нежелани последици върху организацията, например идентифициране на възможните опасности и препоръчване на смекчаващи стратегии далеч във времето [215].

Петте типа анализ имат различни цели и решават различен тип задачи. Обикновено те се използват в описаната последователност с цел подобряване на процесите на информирано вземане на решения, както е показано на Фиг. 2.3.



Фиг. 2.3. Приложение на видовете анализ [184]

Според Perumal [180] функциите на специалиста по данни (*Data Scientist*) и анализатора на данни (*Data Analyst*) много си приличат, но специалистът по данни извършва прогностичен анализ на големи и комплексни масиви от данни, които се генерират и обработват в реално време, с цел подпомагане на организацията във вземането на ефективни решения [130]. Затова специалистът по данни трябва добре да борави с методите, основаващи се на теорията на изкуствения интелект, машинното обучение, а също така и със статистическите методи, като анализ на времеви редове, регресия и др., които се използват за прогностично моделиране. Анализът на данни се извършва по-скоро чрез методите на статистическото и математическото моделиране, приложени върху по-малки масиви от данни [43]. В Таблица 2.2 са посочени основните различия между двата професионални профила.

Таблица 2.2. Диференциация между професиите „специалист по данни“ и „анализатор на данни“

Специалист по данни	Анализатор на данни
Изграждане на статистически модели и вземане на решения, базирани на наличните данни.	Писане на персонализирани заявки в търсене на отговор на сложни бизнес въпроси.
Провеждане на експерименти за откриване на причинно-следствени връзки, които обосновават наблюдавано явление. Обикновено се използва А/В тест или епидемиологичен подход към проблема.	Създаване и внедряване на нови показатели за разкриване на доскоро непознати части на бизнеса и/или продукта.
Идентифициране на възможностите за създаване на нови продукти и функции, които данните предоставят.	Решаване на проблеми, свързани с качеството на данните (като пропуски в данните или пристрастия при събирането на данни), и работа с останалата част от инженеринга за постепенно събиране на нови данни.

В научната литература се забелязва нарастване на броя на публикации, свързани с професионалния профил на специалиста по данни и необходимостта от развиване на такъв вид специалисти в организацията. Рамката на притежаваните от тях знания, умения и компетенции се променя и развива през годините. Това се дължи както на развитието на технологиите и феномена „големи данни“, така и на мултидисциплинарната същност на професията.

Търсенето на специалисти по данни расте през последните години, а Davenport и Patil [73] определят тази професия като най-привлекателната за 21. век.

С оглед на гореизложеното очакването към специалиста по данни е да бъде професионалист с много квалификации, както е показано на Фиг. 2.4.



Фиг. 2.4. Умения на специалиста по данни

Някои автори намират това за нереалистично, дори Geringer нарича този вид специалисти „несъществуващи еднорози“ [106]. Други автори смятат, че поради своята специфика науката за данните предполага екипна работа. Според Gartner (2012) екипът трябва да се състои от няколко специалисти: дигитален архивист (*Digital Archivist*), информационен мениджър (*Information Manager*), дизайнери на визуализация на данни и информация (*Data and Information Visualization Designers*), експерт по ИТ право (*Legal IT Professional*) и управител на данни/информация (*Data/Information Steward*) [35]. Според Samuel Greengard (2014) към този списък трябва да се добави експерт по социални науки (*Social Scientist*). Успехът на такъв екип зависи до голяма степен от наличието на експерт, който може да обедини експертизата на различните гледни точки и да осигури цялостен резултат. Тези виждания се споделят от Christozov et al., определящи *Data Science* като трансдисциплинарна област [56].

В литературата съществуват много и различни схващания относно обхвата на областите на умения и компетенции на специалиста по данни. Някои автори твърдят, че специалистът по данни се нуждае от знания за проблемната област, в която провежда изследването. В резултат от това се появи терминът *domain expert*, който възникна, за да коригира тенден-

цията „анализът на данни да се третира като строго техническо действие“ [34]. В действителност след първоначалното събиране на данни е необходимо задълбочаване на знанията за естеството на данните и техните характеристики. Не се очаква специалистът по данни да бъде експерт в конкретната област на изследване, но той трябва да умее да формулира правилните въпроси, да притежава любознателност, аналитично и критично мислене. Ролята му е свързана със създаването от наличните данни на продукти, които придобиват своята стойност от самите данни [148], а крайният продукт да генерира добавена стойност.

Въпреки че в много случаи, когато се оценяват основните компетенции на специалиста по данни, фокусът е върху техническите му умения, работата с данни е многостранна и изисква овладяване и на друг тип умения и способности. Всъщност, за да бъдат използвани ефективно съответните данни, е необходима комбинация от аналитични, статистически, алгоритмични, инженерни и технически компетенции.

В Таблица 2.3. е представен литературен преглед, обхващащ на уменията на специалиста по данни.

Таблица 2.3. Обхват на уменията на специалиста по данни

Автор	Области от профила на специалиста по данни
Cleveland, 2001	Уменията, използвани от специалиста по данни, разширяват областта на статистиката, навлизайки в науката на данните [59].
Press, 2013	Специалистът по данни използва едновременно статистика и компютърни науки [185].
Dhar, 2013	Науката за данните изисква интегриран набор от умения, обхващащ математика, машинно обучение, изкуствен интелект, статистика, бази данни и оптимизация, както и deep learning. [80]
Hayes, 2014	Специалистът по данни използва инструменти и методи от статистиката и машинното обучение [244].
Granville, 2014	Специалистите по данни не са статистици, нито анализатори на данни, нито компютърни специалисти, нито софтуерни инженери, нито бизнес аналитици. Те имат познания във всяка от тези области, както и в някои области извън изброените [109].
Ayankoya, 2014	Обхватът на уменията, изисквани от специалиста по данни, включва три основни области: компютърни науки, статистика, както и знания за конкретната предметна област [29].
Ismail, 2016	Необходимите умения на специалиста по данни се свеждат до пет основни: бизнес, статистика, машинно обучение, комуникация и анализ. Визуализацията и комуникационните умения позволяват да бъдат интерпретирани данните [130].
Costa and Santos, 2017a, b	Базата от знания, очаквани от специалиста по данни, надхвърля знанията му по компютърни науки и статистика, а също и връзката между тях [64, 65].

Таблица 2.4 обобщава мненията на изследователите за уменията, включени в професионалния профил на специалиста по данни.

Таблица 2.4. Умения, прилагани от специалиста по данни – литературен преглед

Автор	Специалистът по данни:
Bharadwaj, 2000	Притежава умения за работа с ИКТ, но комбинирането на <i>техническите</i> с <i>меки</i> умения е много важен фактор при извличането и управлението на знанията в организацията [36].
Davenport and Patil, 2012	Притежава уменията да <i>извлича отговорите</i> на важни бизнес въпроси, като се възползва от наличните неструктурирани данни [73].
Viaene, 2013	Прилага <i>научни методи</i> за откриване на знания и модели на данните [228].
Shum et al., 2013	Има способност за аналитично мислене и решаване на проблеми [203].
Mohanty, 2013	Практик е, който събира необработени данни и използва своите умения, знания и опит в областта на анализа, за да ги преобразува в полезна информация [210].
Hayes, 2014	Знае как да извлече смисъл от данните и да ги интерпретира, като използва инструменти и методи от статистиката и машинното обучение [244].
Van der Aalst, 2014	Инженер е, който има количествени и технически умения, креативен и комуникативен е и е в състояние да реализира цялостни решения [224].
Сао, 2014	Има уменията да получава знания от хетерогенни източници в реално време, като гарантира прозрачност и сигурност на моделите и резултатите от анализа [47].
Burtch, 2014	Трябва да има както технически (Analytics, SAS, R, Python, Coding, Hadoop, SQL, Database и др.), така и нетехнически умения (любознателност, находчивост, проникателност, комуникационни способности и др.).
Manieri et al., 2015	Експерт е, който манипулира с данни и извлича полезни знания от тях, като осигурява добавена стойност за организацията [151].
Sicular, 2015	Необходим е в рамките на различни организации, което прави трудно предоставянето на пълен и последователен списък с необходимите му умения. Профилът му изисква умения за съхранение и анализ на данни, умения за преобразуване и комуникация [206].
Gehl, 2015	Притежава познания относно бизнес стратегията и функционирането на организацията [105].

Автор	Специалистът по данни:
Christozov and Toleva-Stoimenova, 2015	Притежава критично мислене, което прилага при проверка на получената информация, източниците и ограниченията на приложните технологии. Критичността е фактор за генериране на полезни знания от придобитата информация [55].
Cheryl et al., 2015	Притежава умения за работа с традиционните системи за управление на релационни бази данни, а също и способности за извличане, преобразуване, зареждане и извличане на данни [20].
De Veaux et al., 2016	Притежава умения за получаване, wrangling, куриране, управление, обработка, изследване на данни, дефиниране на въпроси, извършване на анализи и комуникация.
Ismail, 2016	Използва нетрадиционен аналитичен подход, софтуер и инструменти за визуализация на данни, за да проучи поведението на данните [130].
Suhailis, 2016	Притежава умения за моделиране, анализ и обработка на данните, статистика, бизнес умения, меки умения и технически умения.
Christozov et al., 2016	Притежава трите основни категории компетенции при работа с големи данни: (1) възможност за извличане на полезни данни от огромни и разнообразни хранилища; (2) възможност за проверка на получените данни; (3) способност за интерпретация на получените данни в контекста на проблема и извличане на полезни модели и връзки [54].
Costa and Santos, 2017	Трябва да има силни социални и лични способности (комуникация, бизнес проникателност и любопитство) [64, 65].
Parks, 2017	Трябва да притежава интуиция, за да осмисля бързо данните и да определя важните им аспекти, като пренебрегва нерелевантната информация. Уменията, необходими на специалиста по данни, може да бъдат групирани в пет ключови области (функции за анализ на данни, програмиране и инструменти, техники и решения за машинно обучение, интердисциплинарни знания, способност за критично мислене) и общо 55 компетенции [177].

След анализ на описаните в Таблица 2.3 и Таблица 2.4 схващания за знанията и уменията на специалиста по данни може да се направят следните изводи:

- Голяма част от изследователите са на мнение, че знания и умения от области като статистика, математика, програмиране, както и интердисциплинарни знания, са задължителни при обработката и анализа на данни.
- Съществуват значителни различия по отношение на програмните инструменти, които стоят в основата на професионалния профил на специалиста по данни. Например инструменти като *Spark*, *Radiant*, *Tensorflow*, *Tableau*, *Qlikview* и *SAS* се откриват рядко в научната литература. Възможно е тези инструменти да са нови на работното

място и да не са напълно възприети в академичните среди [177]. Освен това развитието на инструментите е много динамично, а от по-съществено значение и с по-дълъг времеви хоризонт е предлаганата от дадения инструмент функционалност.

- По отношение на техниките и решенията за машинно обучение авторите обикновено препоръчват използване на минималните изисквания в съответствие с принципа, известен като „бръснача на Окам“.

- Умението за практическо приложение на дървета на решения, неврални мрежи, вектори и др. се оказват съществени за профила на съвременния специалист по данни [177].

- В представените изследвания не са засегнати ключови атрибути, като способност да се зададе интелигентно въпрос при дефинирането на проблем, и критично мислене. Често авторите поставят тези способности в категорията на меките умения и ги свързват със способностите за комуникация, визуализация на данни и експертни познания.

2.2.2. Рамка на ключовите компетентности

С цел идентифициране на ключови компетентности и изграждане на обобщаващ професионален профил на специалиста по данни ние ще разграничаваме три обобщаващи категории базови умения, а именно:

- твърди умения;
- меки умения;
- аналитични умения.

Твърдите умения са технически умения, които включват боравене с езици за програмиране (*Python*, *R*, *Java*, *Ruby*, *Clojure*, *Matlab*, *Pig* и *SQL*), базови знания по математика и статистика, като машинно обучение, концептуално моделиране, статистически анализи, прогнозно моделиране и тестване на хипотези и др.

Категорията **меки умения** включва психологически и емоционални компетентности, които помагат на хората да се справят ефективно с предизвикателствата в личния и професионалния си живот. Такива са организационно-управленските умения, свързани с вземане на решения, решаване на проблеми, ефективна комуникация, междуличностни отношения и други „нетехнически“ умения [130]. За разлика от твърдите умения меките умения не са специфични за дадена сфера.

Едновременно с меките и твърдите умения [29] специалистът по данни трябва да прави оценки, прогнози и препоръки за развитие на бизнес процесите, което е свързано с умения за предиктивен анализ, визуализация и моделиране на данни и машинно обучение. Очевидна е връзката между аналитичните и твърдите умения, тъй като процеси-

те на вземане на решения, изготвяне на стратегии и осъществяване на експериментални изследвания се базират на данни, които се получават от други данни на базата на някаква предварителна обработка. Специалистите по данни използват своите аналитични умения, като оперират свободно с притежаваните от тях твърди умения. От друга страна, областите на аналитичните и меките умения също имат обща част, която е свързана с критично и аналитично мислене, креативност, способност за решаване на проблеми и др. **Аналитичните умения** ще се разглеждат като отделна категория, за да се подчертаят спецификата на работата на специалистите по обработка и анализ на данни и ролята им за постигането на добри резултати.

Моделът на трите ключови категории умения, които трябва да притежава специалистът по (големи) данни, е представен в Таблица 2.5.

Таблица 2.5. Компетентност и компетенции в науката за данните

	Меки умения	Твърди умения	Аналитични умения (меки и твърди)
Задачи	Създава и продава информация, основана на данни, устно и визуално.	Осигурява качество на данните. Изгражда статистически модели. Създава продукти/ платформи от данни. Създава визуализация на данните. Интегрира данни от множество източници независимо от тяхната структура и обем.	Идентифицира богати източници на данни. * Анализира очаквана стойност. * Взема инженерни ефективни решения. * Намира оповори на важни бизнес въпроси. * Подобрява вземането на решения. * Предлага нови бизнес направления. * Прилага аналитично мислене. * Използва и анализира данни. * Очертава причинно-следствени заключения.

	Меки умения	Твърди умения	Аналитични умения (меки и твърди)
Умения	Бизнес проникателност; комуникационни умения; общуване; предприемачество; любопитство	Компютърни науки; изкуствен интелект; автоматизирани анализи на данни; статистика; големи данни; бази данни; машинно обучение; математика; мрежи; програмиране; облачни изчисления; разпределени изчисления; обработка на данни; погълчане на данни; извличане на данни; подготовка на данни; инструменти за данни	Академични изследвания; формулировка; интердисциплинарност; научни и изследователски методи; дизайн, анализ и интерпретация на данни; визуализация на данни; хранилища за данни
Компетенции	Да разбира основните бизнес цели и стратегии, тъй като това ще позволи максимално улъбняване на получените знания от данните. Да умее да разбира заинтересованите страни и да подпомага процеса на вземане на решения. Да умее да комуникира и разпространява констатациите.	Да притежава технически умения за статистическа обработка на данни, които да прилага при проектиране и интерпретиране на експерименти и изграждане на модели и прогнозиране. Да умее да създава артефакти за данни или да оптимизира съществуващите.	Да познава методи за анализ на данните, автоматизиращи изграждането на аналитични модели. Да подпомага управлението на бизнеса чрез подобряване на ефективността на процесите на вземане на решения.

В Таблица 2.6, на базата на модела „твърди – меки – аналитични умения“, са систематизирани основните категории, задачи и умения на специалистите по данни [194, 195].

Голяма част от ИТ специалистите днес притежават умения да боравят с малки по обем данни. Работата с големи данни обаче изисква повече от техническа грамотност, статистика, математика, програмиране и работа с база данни. Липсата на меки умения, както и на аналитично мислене, може да се окажат основни бариери за формирането на знание от големите данни.

Въз основа на създадената рамка от умения и спецификата на трите основни категории умения може да се направят следните изводи:

- твърдите умения може да се разглеждат като основна характеристика на ИТ специалистите;
- меките умения са ключови умения за информационните медиатори (т.нар. инфоброкери);

- аналитичните умения са основна категория умения за анализаторите на данни и учените.

Таблица 2.6. Основни задачи и необходими умения в науката за данните

Категория	Задачи	Умения
Извличане	Подбира, класифицира, събира, сравнява, конфигурира, противопоставя, дефинира, демонстрира, описва, изгълнява, обяснява, намира, идентифицира, илюстрира, етикира, показва, търси съвпадение, именува, пропуска, експлоатира, очертава, припомня, перифразира, показва, резюмира, разказва, превежда.	Твърди + меки
Проверка	Прилага, анализира, изгражда, конструира, разработва, изследва, експериментира, идентифицира, намира, инспектира, моделира, мотивира, организира, избира, опростява, разрешава, проучва, тества, визуализира.	Твърди + меки + аналитични
Интерпретация	Адаптира, оценява, променя, комбинира, компилира, композира, обобщава, критикува, създава, решава, защитава, проектира, дискутира, определя, опровергава, оценява, усъвършенства, влияе, избобретява, отсъжда, оправдава, оптимизира, планира, предвижда, приоритизира, доказва, препоръчва, решава.	Твърди + меки + аналитични

На Фиг. 2.5 е показана диаграма на Вен за уменията на трите вида специалисти.



Фиг. 2.5. Профил на основните умения на специалиста по данни

Твърдите, в комбинация с меките умения, при наличието на аналитично мислене и умения за провеждане на анализи, формират общата компетентностна рамка за съвременния специалист по данни.

Забелязват се следните зависимости:

1. Наличието на твърди умения е база за извличане на данни и формиране на нови знания от тях.

2. Без наличието на аналитични умения извлечените данни са просто преобразувани данни, които нямат информационна стойност и от тях не може да се придобие полезно за организацията знание с цел информирано вземане на решения.

3. Познаването на бизнес плана и основните насоки за развитие на организацията е важна предпоставка за осъществяване на правилен алгоритъм при изследване на данните и постигане на търсените резултати.

Представеният модел на компетентностите ни дава основание да твърдим още, че дигиталното разделение в контекста на големите данни на фирмено равнище се увеличава догълнително поради грешния подбор на човешки капитал при назначаването на специалисти по данни. Уменията за управление на база данни, изготвяне на статистики и програмни езици са само част от нужните умения при работа с големи данни. Специалистът по данни трябва да притежава експертиза едновременно в три категории: меки, твърди и аналитични умения. Тези категории са взаимосвързани и взаимнозависими. Днес наличието им е задължително за постигането на успех при работа с (големи) данни.

2.2.3. Аналитични умения

Както стана ясно, специалистите по данни трябва да разполагат с широк спектър от компетенции в различни области, които разделихме в три основни категории – твърди, меки и аналитични. В тази подточка ще разгледаме подробно аналитичните умения на специалистите по данни, които според нас са много важен елемент от техния профил предвид спецификата на работата им.

През последните десетилетия се наблюдава преминаване от анализ на малки и прости данни и тестване на хипотези към анализ на големи комплексни данни. Процесът на обработка и анализ на големите данни се характеризира с редица проблеми и предизвикателства, пред които се изправят специалистите по данни. Според Microsoft Azure's blog [160, 161] те трябва да отговарят на пет типа въпроси:

1. Колко? (Регресия)
2. Коя категория? (Класификация)

3. Коя група? (Клъстеризация)
4. Нормално ли е? (Откриване на аномалии)
5. Какво да изберем? (Препоръка)

Анализът на данни е сложен многостъпков процес, който включва разбиране на проблемната област, събиране на данни, почистване и предварителна обработка, интеграция, разпознаване на шаблони, анализ, интерпретация и тълкуване на резултатите. Както при производството на всяка услуга или артефакт, цената, навремето и качеството определят успеха на анализа [190].

Sternberg and Grigorenko разделят мисловните умения в три категории: аналитично, креативно и практическо мислене, като аналитичното мислене включва процесите на решаване на проблеми, основани на знанието, и на вземане на решения [212]. Robbins твърди, че както аналитичното, така и логическото мислене са необходими при решаването на проблеми [198]. Аналитичното и критичното мислене често се свързват с *direct thinking* (мислене, което е пряко насочено към решаване на проблема, търсене на истината и разбиране с фокус върху желания резултат [66, 67]). В литературата уменията за критично, логическо и аналитично мислене често са свързани, а понякога се смесват и отчасти се припокриват [192].



Фиг. 2.6. Нетехнически аналитични умения

Ще разглеждаме аналитичното мислене като по-общо понятие, което обхваща елементи от логическото и критичното мислене. Този вид мислене и умения са изключително ценни в дейности като анализ на факти, аргументи, твърдения, достигане до заключения на базата

на дедуктивни и индуктивни разсъждения, оценка на резултати, вземане на решения и решаване на сложни проблеми [97, 237, 90, 91], които са тясно свързани с работата на специалистите по данни. Ще приемем, че те са свързани със и формират аналитичните умения на специалиста по данни.

Аналитичните умения и компетенции започват да се изграждат още в семейството, като на всеки етап от образованието се разширяват и подобряват. Основните дисциплини, които подпомагат това, са математика, статистика, програмиране и др.

Овладеяването на компетенцията за аналитично мислене изисква развиване на набор от умения, като:

- разпознаване в разбъркани описания на дадена ситуация, на съдържащите се в тях информативни елементи и класифицирането им до степен на приемлива надеждност или сигурност;
- приемане на наличието на празноти в информацията, за запълването на които са нужни предварителни теоретични модели. Липсата на достатъчно или на достатъчно надеждна информация може да доведе до търсене на нова информация, до издигане на разумни хипотези или до оценяване, преди да бъдат направени заключения;
- използване на аналитични инструменти за организиране на наличната информация и показване на връзките между различните компоненти (визуализация за описване на данни по смислен, структуриран начин, като фигури, таблици, карти на концепции, графики и др.) и за извличане на знание от данни чрез използване на статистически и други техники.

В компетентностния модел, изложен в предходната точка, са посочени следните задачи пред специалиста по данни, които са свързани с неговите аналитични умения [194]:

- сравняване, класифициране и подреждане в последователност на абстрактна материя, като данни и информация;
- търсене и откриване на причинно-следствени връзки;
- формулиране на хипотези;
- прилагане на дедуктивно и индуктивно мислене;
- прогнозиране и планиране;
- разглеждане, сравняване и критична оценка на алтернативите;
- генериране на нови идеи;
- създаване на оригинални и иновативни решения;
- модифициране на съществуващи подходи;
- запис и визуализация на сложни решения.

Това ни дава основание да твърдим, че аналитичните умения трябва да бъдат включени в профила на специалиста по данни като отделна категория. Очевидно аналитичното, критичното и логическото мислене са важни фактори, които определят тези ключови за професионалиста умения. От друга страна, развитието на аналитичните умения може да се разглежда като възможност за преодоляване на появилото се през последните години разделение между лица, притежаващи умения да „четат“ резултатите от обработката на големите данни, и останалите, които използват услугите на медиатор, за да се възползват от наличните данни. Във връзка с това е редно да се направи разделение между аналитичните умения, които са зависими от твърдите умения, и онези, които са част от човешките мисловни способности.

Тук ще направим класификация (таксономия) на аналитичните умения и компетенции, необходими за анализ на (големи) данни. Тази таксономия може да се използва като ориентир при проектиране и изпълнение на различни видове програми, квалификации и/или степени за обучение в областта на науката за данните. Ние следваме до голяма степен публикациите на Lumina Foundation [23] по отношение на *Degree Qualification Profile (DQP)*, като ги съобразяваме с нашия собствен опит.

Общата рамка, използвана тук, включва три категории променливи, описани от събраните данни, които са предмет на анализ:

- контролирани променливи (*Controlled variables*);
- неконтролирани променливи (*Uncontrolled variables*);
- резултати (*Outcomes*).

Първите две категории се наричат още фактори, предиктори, входни данни. Основна цел на анализа на данните е да се намери модел, който да описва връзката между тези три категории променливи. Наличието на аналитични компетенции позволява на специалистите по данни да постигнат правилно разбиране за „причините и последиците“, ръководещи развитието на определени обстоятелства, и превръщането им в полезна форма.

Ще разгледаме процеса на анализ на данни, като класифицираме необходимите аналитични компетенции в отделните етапи по следния начин:

- A. идентифициране и анализиране на същността на проблема;
- B. идентифициране и оценка на информационните потребности, качеството на данните и степента на релевантност на данните спрямо проблема;
- C. синтезиране на решение.

А. Идентифициране и анализиране на същността на проблема

Началният етап е свързан с разбирането на проблемната област, дефинирането на целите на изследванията и съответните изисквания от гледна точка на потребителя. Тази категория компетенции се отнася до способността да се анализират конкретни обстоятелства, да се поставят факти в даден контекст, да се идентифицира съществуването на проблема и той да се формулира. Правилното дефиниране и структуриране на проблема са от решаващо значение за постигането на релевантни и приложими резултати.

Идентифициране на проблема

Наличието на проблем се установява чрез извършване на сравнение между очакваните резултати и данните, описващи реалното състояние. Допустими са колебания в дадени граници, на базата на случайността на неконтролираните фактори.

Това е процес на установяване на това дали разликата между очакваното и постигнатото се основава на случайни обстоятелства, или не. Последното показва наличието на проблем и дефинира целта на изследването. Специалистът по данни трябва да изясни проблема и да ограничи обхвата на проучването до разбираеми, прости, кратки и изпълними задачи, които може да бъдат формулирани като въпроси.

Идентифициране на причините за възникване на проблема

Тези компетенции обхващат аспектите, свързани с разбиране на контекста на възникване на проблема, и дават възможност да се направи оценка на приложимостта на решението.

Основната цел е да се гарантира, че всички заинтересовани страни разбират контекста на изследвания проблем: кой, какво, къде, кога, защо и как [46]:

- **Кой?** – Кой причинява проблема? Кой казва, че това е проблем? Кой (не) са засегнати?
- **Какво?** – Какво се случва, ако проблемът не бъде решен? Какви са симптомите? Какви са въздействията?
- **Къде?** – Къде възниква проблемът? Къде има влияние?
- **Кога?** – Кога възниква този проблем? Кога за първи път започна да се случва?
- **Защо?** – Защо възниква проблемът?
- **Как?** – Как трябва да работи системата? Как се обработва настоящем?

Специалистът по данни трябва да притежава необходимите компетенции, за да отговори на следните въпроси:

- Какви (входящи) фактори са предизвикали проблема?
- Какви са съответните променливи?
- Каква е приемливата променливост на неконтролираните входни данни?

Често се използват различни инструменти за анализ на причини, например причинно-следствената диаграма на Ишикава, или т.нар. „рибена кост“. Чрез нея може да се идентифицират множество възможни причини за даден проблем и да се сортират идеите в полезни категории. Тя демонстрира връзката между ефектите и техните причини и се използва за откриване на основните причини. От своя страна основните причини може да имат подпричини и т.н. на много нива. Други инструменти за анализ на причини са диаграма на Парето, диаграма на разсейване, стратификация, хистограми и др.

На този етап се идентифицират променливите, които трябва да се предскажат, но също така и категорията на приемливите решения за дадената категория проблеми и обхватът на приложимите аналитични методи. Ако се търси прогнозиране на конкретна стойност, естественият избор е използване на регресия. При определяне на принадлежност – класификация, с допълнително отчитане дали става дума за обучение с учител, или без учител, в зависимост от наличните данни.

Б. Декомпозиране на проблема

Следващата стъпка след идентифицирането на проблема и дефинирането на задачата е декомпозицията. Поставената задача се разбива на подзадачи, всяка от които е със собствени данни, цели, изчисления и резултати. В повечето случаи съществуват повече от един варианти за декомпозиране на задачата, които обикновено водят до различно решение.

Тези задачи изискват задълбочено разбиране на проблема и взаимовръзките между променливите.

Оценяване на това дали факторите, които влияят върху решението на проблема, може да бъдат контролирани

Причинно-следствената диаграма помага на експертите да изследват основните променливи и техния характер, да определят дали са измерими, може ли да бъдат контролирани, или зависят от околната среда и случайни фактори, и т.н. Независимите (контролируеми) променливи, наричани още „предиктори“, се използват за прогнозиране на стойностите на зависимите променливи.

В. Идентифициране и оценка на информационните потребности, качеството на данните и релевантността на данните спрямо проблема

На този етап е необходимо да се направят критичен анализ и изследване на данните с цел задълбочаване на знанията на анализатора за тяхното естество. Също така трябва да бъдат идентифицирани проблеми, свързани с качеството на данните, да се получи първоначално мнение за характера на данните, да се намерят интересните подмножества на данните, за да бъдат формирани първоначални хипотези.

Тази категория компетенции включва също способността за анализ на достоверността на изследваните данни, както и компетенции за избор на аналитични инструменти (техники, методи), които са подходящи за конкретния проблем и свойствата на изследваните данни.

Компетенциите, които принадлежат на тази категория, често се наричат „информационни“ компетенции.

Идентифициране и оценка на необходимата информация

Разбирането на проблемната област и контекста, в който възниква проблемът, определя как ще се оценява релевантността на данните. От друга страна, идентифицирането на входните променливи, независимо дали са контролируеми или не, е от критично значение за по-нататъшния анализ и търсене на решение.

Къде може да се намери подходяща информация (източници)?

Дали наличните данни осигуряват достатъчен ресурс за изграждане на модела с приемливо ниво на сигурност? Възможно ли е да се обогати базата от данни с цел подобряване на анализа?

Идентифицирането на по-голямо разнообразие от външни и вътрешни източници на данни може да увеличи вероятността за намиране на скрити модели и корелации, особено когато не е ясно какво точно се търси.

Възможност за достъп, извличане и структуриране на съответните данни

Тази подкатегория обхваща техническите компетенции, необходими за получаване и представяне на данните във вид, подходящ за по-нататъшна обработка. Те включват:

- достъп до и търсене в много различни бази данни и източници на данни (*RDBMS*, *NoSQL*, *NewSQL*);
- интегриране на данните в подходящи хранилища, където може да бъдат обогатявани и анализирани (например *OLAP*, складове за данни, езера за данни);

- структуриране на данните в правилен формат, тъй като данните, получени от различни източници, имат различен формат и размер.

Едно от основните предизвикателства, пред които са изправени изследователите, е как да интегрират всички данни от различни източници, за да извлекат максимална добавена стойност.

Възможност за изчистване и изследване на данните (фаза на трансформиране)

Фазата на трансформиране включва почистване и премахване на шум от данните (премахване на несъответствия, справяне с липсващи данни и т.н.), нормализиране, агрегиране и обобщение. Целта е да се получат ефективно организирани данни, подготвени за същинските анализи и извличане на знания.

В същото време при работа с големи данни с обем, измерващ се в терабайтове, автоматичното почистване на данните може да доведе до загуба на потенциално полезна информация, а ръчното е практически невъзможно. Умението да се намери компромисно решение, е ключово и се определя от разбирането за същността, произхода и пътя на събиране и запис на данните.

Етапът на трансформиране следва етапа на проучване на данните. Това е като мозъчната атака при анализа на данни. Той включва анализ на данните, намиране на зависимости и изместване (bias), откриване на модели и зависимости сред суровите данни за по-дълбоко разбиране и повишаване на увереността, че се познават ограниченията и качеството на събраната информация. Функционалният инженеринг е съществена част от изграждането на всяка интелигентна система. Той се занимава с подбор на функции (съкращаване на функциите, които добавят шум в информацията) и изграждане на функции (създаване на нови функции). Функционалният инженеринг е важна част от подходите при почистване на данните, тъй като значително намалява шума, решава проблема с обема на извадката и с нейната многомерност и дори може да реши проблемите с редки данни (outliers).

Г. Способност за проверка и оценка на достоверността на получената информация

Достоверността на данните се влияе от три основни фактора: надеждност на източниците на данни, нормализиране на данните и времето на създаване и регистриране на данните. Данните, събрани от авторитетни научни организации, обикновено се характеризират с висока достоверност, защото изследователите, като цяло, спазват изс-

ледователската етика и следват утвърдени научни практики. Понякога големите данни се събират от търговски компании или от системи, които не са пряко контролирани от надеждна организация и може да съдържат съществени отклонения или неверни стойности (например заради ботове в социалните медии).

Компетенцията за оценка на достоверността на данните е много важна и има решаващо значение за качеството на получения вследствие на анализа резултат.

Д. Способност за оценка на качеството на данните

Качеството на данните се оценява на базата на различни критерии, като *консистентност, точност, прецизност, пълнота, навременност, уникалност* и др. Тази категория обхваща компетенциите, които са необходими за избор на правилните мерки, в конкретния случай – и на умения за оценка дали данните отговарят на очакванията за качество. За откриване и коригиране на грешки и несъответствия в набора от данни с цел подобряване на качеството се използва различен набор от умения и инструменти. Доброто познаване на архитектурата, технологията, корпоративната култура и други фактори често е особено важно.

Техническите умения са задължителни на този етап от работата с данни, включително за съпоставяне, стандартизиране, свързване на записи, съвпадение, почистване, изчистване на данни, профилиране на данни и одит, мониторинг на данните.

Е. Компетенции за избор на аналитични инструменти

Това са технически компетенции за прилагане на правилните аналитични методи и инструменти в съответствие с наличните данни. Специалистът по данни трябва да умеє да трансформира суровите данни във вид, подходящ за съответните инструменти.

При това анализаторите са изправени пред голямо разнообразие от ограничения: аналитична сложност, скорост, точност и прецизност, размер и сложност на данните. Налага се да се правят редица компромиси, за да се постигне баланс.

Ж. Синтезиране на решение

Тази категория компетенции е свързана със способността за ефективно използване на инструментите за анализ на данни. Изискват се специфични умения по математика, статистика, изкуствен интелект, както и ефективно използване на компютърните технологии. Често тези компетенции като цяло се определят като компетенции по „машинно обучение“.

Разбиране на ограниченията на приложен аналитичен инструмент

Истината е, че не съществува „едно-единствено решение“ или „най-добър алгоритъм“. Няма алгоритъм или решение, които да са универсално приложими за всички ситуации – изборът на най-добрия алгоритъм зависи от проблема, размера, наличните ресурси и др.

Повишеният капацитет за обработка на големи данни създава тенденция към включване на нерелевантни данни. От друга страна, използваните аналитични техники като „черна кутия“ може да подвеждат анализаторите при интерпретирането на резултатите в контекста на проблемната област. Вземането на решения при използването на „черна кутия“ е сериозно ограничение, тъй като много изисквания за изпълнение на политиката на управление се нуждаят от ясни обяснения на решенията (например да се обясни на клиента защо трансакцията му е блокирана).

Големият обем данни увеличава вероятността файловете с данни и документите да съдържат лична и чувствителна информация. Тя може да засяга неприкосновеността на личния живот и сигурността на хората, а това води до някои предизвикателства относно етичните аспекти на анализа на данни. Понякога при превода на културни клишета и стереотипи се въвежда факторът субективност. При някои обстоятелства дори показването на различни реклами може да означава, че потребителите на тази услуга се профилират по начин, в който се открива дискриминация, например въз основа на раса, пол, политически убеждения, религия и т.н. Специалистите по данни трябва да се справят с подобни проблеми и да не разчитат на изцяло автоматизирано профилиране.

Формулиране на хипотези

Компетенциите за формулиране на предположения при прилагането на аналитични техники са от съществено значение, за да се даде възможност за полезно тълкуване и обосновка на получените резултати. Специалистите по данни трябва да могат да генерират хипотези за начините, по които съответната система може да се държи, ако се промени по определен начин. Ясно е, че атрибутите, които не са в модела, няма да имат ефект върху прогнозите на модела [134].

Генерирането на хипотези позволява:

- експеримент с теориите за данните;
- възприемане на системно мислещия подход към проблема, който трябва да бъде решен;
- изграждане на по-сложни модели, основани на предишни хипотези и разбиране.

Сравняване на различни приложими възможности за действие

Изборът на инструмент, повишаващ информираността за даден казус, изисква способност да бъде осъществявано сравнение между наличните/съществуващите инструменти въз основа на фактори като: (1) свойства на наличните данни и (2) цели на анализа им.

Специалистът по данни трябва да избере измежду методите, използващи математическа статистика и машинно обучение, в зависимост от естеството на данните, проблема/случая, за който се прилагат, и наличните изчислителни ресурси. В повечето случаи специалистите използват широк спектър от технологии, за да имат възможност да избират и тестват многобройни методологии за моделиране и да изследват прогнозните резултати.

Моделирането е мястото, където изкуството среща науката. Изкуството идва от априорната интуиция и предположения за връзката между променливите. Науката идва от провеждането на количествен анализ на данните и използването на голяма колекция от модели. Подборът на модели включва ограничаване на броя на параметрите и избор на модел, който е разбираем и приложим от заинтересованите страни [197].

Въпросите, свързани с тази компетенция, са [160, 161]:

- Изглежда ли моделът валиден, приложим и релевантен спрямо набора от данни?
- Полезен ли е моделът (поведението на модела) за експертите в областта?
- Имат ли смисъл стойностите на параметрите в контекста на домейна?
- Достатъчно точен ли е моделът, за да отговори на целта?
- Съществуват ли допустими или недопустими грешки?
- Разполагаме ли с достатъчно данни, или имаме нужда от повече информация?
- Нужни ли са трансформации или корекции на данните?
- Поддържа ли се избраният модел по време на жизнения цикъл на средата, в която функционира?
- Необходима ли е различна форма на модела за решаване на бизнес проблема?

Проверка на получения модел

Тази категория включва компетенции за прилагане на техники за проверка на приложимостта на получения модел (*cross validation and receiver operating characteristic (ROC) curve analysis, etc.*). Специалистът по дан-

ни може да прецени доколко моделът е ефективен, като го приложи към някои от данните за обучението и сравни прогнозата с известна стойност. Обикновено моделът се тества в по-голям набор от случаи, за да се определи неговата приложимост.

Представяне на резултатите

След извършването на анализ и извличането на полезната информация специалистът по данни трябва да представи информацията по подходящ начин на заинтересованите страни, така, че те да могат да я разберат и интерпретират правилно и да я използват ефективно. Това предизвикателство изисква от специалистите по данни да могат да боравят с различните инструменти за визуализиране на данни, както и с техниките и технологиите, използвани за създаване на изображения, диаграми, анимации и др. В общия случай методите за визуализация зависят от използвания модел за машинно обучение и други специфични характеристики на процеса на работа на специалистите по данни.

За постигане на ефективна визуализация на извлеченото знание често се съчетават разнородни области, като комуникация, психология, статистика и дори изкуство [30].

Освен разгледаната класификация на аналитичните умения на специалистите по данни може да се предложат и други класификации.

Алтернативни класификации са направени в глава I по отношение както на използваните технологии, така и на проблемната област, в която протича процесът на обработка и анализ на данните.

В глава I може да се разглежда класификация на аналитичните умения в зависимост от използваните технологии за анализ и извличане на знание.

Трета класификация може да бъде направена в зависимост от областта, в която се прилагат аналитичните умения. Спецификата на проблемната област, в която се генерират и събират (големите) данни, определя необходимия тип архитектура за съхраняване, обработване и извършване на анализ на данните. Има няколко начина за характеризиране на данните [37]:

Според времевия интервал, в който трябва да се извършва анализ:

- реално време (финансови потоци, обработка на сложни събития, установяване на неоторизиран достъп, откриване на измами);
- близко до реалното време;

- партида (търговия на дребно, съдебна медицина, биоинформатика, геоданни, исторически данни от различен тип).

Според степента на структурираност на данните:

- структурирани (търговия на дребно, финансови, биоинформатика, геоданни);
- полуструктурирани (уебдневници, електронна поща, документи);
- неструктурирани (изображения, видео, данни от сензори, уебстраници).

Според сектора, който генерира данни и се нуждае от извличане на информация от тях:

- финансови услуги (високочестотна търговия);
- търговия на дребно (анализ на поведението);
- мрежова сигурност (откриване на нарушения, АРТ);
- мащабна наука (биоинформатика, физика с висока енергия);
- социални мрежи (анализ на чувствата, социални графики);
- „интернет на нещата“/сензорни мрежи (климатични промени, откриване на аномалии);
- визуални медии (анализ на терена, анализ на аудио, графики, изображения).

2.2.4. Етични умения

Днес много голяма част от генерираните големи данни, които се обработват от специалистите по данни, съдържат лична и чувствителна информация. Във връзка с възможностите, които предоставят на експертите съвременните аналитични инструменти, възникват редица въпроси, които имат етичен характер [44, 45]:

- Необходимо ли е даден човек или неговите данни да бъдат включени като част от голямото количество данни?
- Може ли да се анализират и изследват нечии данни, в случай че собственикът им не е информиран за това?
- Кой ще поеме отговорност за това?
- Какво е информирано съгласие?

Достъпът до лична информация (включително предпочитания при пазаруване, детайлни записи за обаядания и постове в социалните мрежи) води до увеличаване на проблемите, свързани с неприкосновеността на личния живот [33, 132]. Експертите по данни разполагат с техническа инфраструктура за достъп до огромно разнообразие

от източници на данни и инструменти за анализ и профилиране, докато потребителите не са наясно, че публикуваната от тях в определен контекст информация може да бъде използвана за други цели [45]. Данните могат да бъдат публични (или полупублични), но публикуването им от даден човек в интернет не е равносилно на пълно разрешение за тяхното използване от всички и за всичко. Специалистите, които анализират големи данни, трябва да вземат под внимание това, че има значителна разлика между това да бъдеш обществено достъпен (например да седиш в парка), и да си публична личност (т.е. активно да привличаш вниманието върху себе си) [155].

Влизането в сила на Общия регламент относно защитата на данните (*General Data Protection Regulation, GDPR*) налага необходимостта всяка организация да назначи служител по защита на данните, чиято основна функция е да насърчава и контролира спазването на основните принципи и да защитава правата и свободите на субектите на данни.

Общият регламент относно защитата на данните влезе в сила на 25 май 2018 г. и се отнася за всички организации и предприятия, които обработват лични и маркетингови данни на европейски граждани. *GDPR* въвежда множество значителни промени спрямо действащата правна рамка на Директивата за защита на данните от 1995 г. – разширява обхвата на понятието „лични данни“, въвежда строги изисквания за валидност на съгласието като основание за обработване, а също за прозрачност, отчетност и др. Съществуват шест основни принципа, заложиени в *GDPR*, които гарантират правата и сигурността на личната информация:

1. законосъобразност, добросъвестност и прозрачност;
2. минимум данни;
3. ограничаване на целите;
4. ограничаване на съхранението;
5. цялостност и поверителност;
6. точност на данните;
7. отчетност.

GDPR въвежда и разширява правата на субектите на лични данни до следните осем права: право на достъп до информация, право на корекция, право на заличаване (да бъдеш забравен), право на ограничаване на обработването, право на възражение, право на жалба пред надзорния орган.

Всичко това поставя завишени изисквания към експертите по данни относно защитата на личните данни. Те трябва да извършат информационен одит, да очертаят потоците с данни, да прегледат за-

коновите основания и договорите и да въведат персонални, организационни и технически мерки за защита на личните данни. Тези умения ще диференцираме като отделен тип и ще наричаме „етични умения“.

Ключовите компоненти на Общия регламент относно защитата на данните не са непременно нови. Може да се каже, че винаги е имало етични норми в областта на дигиталната информация. Днес, в епохата на големите данни, *GDPR* ще има голямо влияние върху дейностите на специалистите по данни, доколкото те работят с лична и чувствителна информация [219, 229]. В допълнение, анализите на големи данни обикновено включват събиране и анализиране на възможно най-много данни, а в много случаи – и на всички данни в даден набор, а не извадка от тях („*n* = всички“). В този контекст принципите за ограничаване на целта на Регламента, минимизиране на данните и ограничаване на съхранението трябва да се прилагат през целия жизнен цикъл на личната информация, което създава известни противоречия.

Както стана ясно, от гледна точка на науката за данните спецификата на работата на анализаторите може да се разглежда като процес от пет основни стъпки:

- (1) определяне на целта;
- (2) събиране на данни;
- (3) проучване и откриване на данни;
- (4) моделиране на данни;
- (5) комуникация и визуализиране на резултата.

Очевидно е, че правилата на *GDPR* ще рефлектират най-силно върху стъпки (1), (2) и (5).

Първата стъпка в процеса на работата на анализаторите е идентифицирането на целта. Тя включва също дефиниране на това кои данни ще бъдат събирани и под каква форма (например дейност по обработка на данни за поведението на потребителите). *GDPR* подчертава, че лицата имат право да бъдат информирани преди събирането на данните за тях, както и да получат информация какви лични данни ще бъдат събирани, по какъв начин и с каква цел. Затова изследователят трябва да информира субекта на данните и да поиска неговото съгласие. Съгласно *GDPR* съгласието трябва да бъде информирано, недвузначно, дадено с ясен положителен акт и доказуемо. Прозрачността също така изисква всяка информация и комуникация със субекта на данни да бъде лесно откриваема и достъпна и да се осъществява на ясен език.

Втората стъпка в разглеждания процес е свързана със събирането на данни, което трябва да бъде съобразено с принципите на *GDPR* за минимизиране на данните и ограниченията за тяхното съхранение. Така например, ако администраторът на данни се интересува от информация за предпочитанията на клиента по отношение на конкретен продукт, данни като семейно положение, заплата, адрес на електронна поща или телефонен номер не са му необходими за изследването. Тъй като специалистът по данни е отговорен за защитата на събраните данни (съхранение, всички архивни файлове, всички версии и копия и т.н.), той трябва да е наясно с последиците от принципа за съхранение на данни. Специалистите по данни трябва да имат информация за съхранението на данни и да преценят дали не са изложени на нарушения на поверителността. Както псевдонимизацията, така и анонимизацията се насърчават в *GDPR* и експертите трябва да бъдат подготвени да използват правилно механизми, които ги правят възможни.

Анонимизацията на данните означава, че тя необратимо унищожава всеки начин за идентифициране на субекта на данни. Получените данни не трябва да може да се свържат с конкретен индивид, с други данни за индивида, нито да бъдат използвани за извеждане на самоличността му. От друга страна, псевдонимизацията е процес на заместване на данни, които директно идентифицират конкретен индивид, с данни, които индиректно го идентифицират (изкуствени идентификатори или псевдоними). Псевдонимизацията не премахва цялата идентифицираща информация от данните, а само намалява връзката с оригиналната идентичност на индивида (например чрез схема за криптиране). Псевдонимните данни са защитени срещу идентификация, но те все още са лични и позволяват повторно идентифициране, докато анонимните данни не може да бъдат повторно идентифицирани.

Петата стъпка от процеса на работата на анализаторите включва комуникация и визуализация, последвани от реализация на резултатите. Принципите на добросъвестно и прозрачно обработване изискват субектът на данни да бъде информиран за съществуването на операция по обработването на данни и да декларира съгласието си за това. Освен това е необходимо субектът да бъде запознат с целите и основанията, които налагат тази обработка, и с контекста на профилирането. *GDPR* включва разпоредби, отнасящи се конкретно до профилирането, което е определено в чл. 4 като: „всяка форма на автоматизирана обработка на лични данни, състояща се от използването на

тези данни за оценка на някои лични аспекти, свързани с физическо лице, по-специално за анализиране или прогнозиране на аспектите, свързани с работата на физическото лице на работното място, икономическата ситуация, здравето, личните предпочитания, ограничения, свързани с надеждността, поведението, местоположението или движенията“. GDPR въвежда ограничения за начина, по който популярните инструменти за профилиране и технологии проследяват на потребителите (като „бисквитките“) и подходите при профилиране. Регламентът не възпрепятства автоматичното вземане на решения или профилирането, но дава на хората квалифицирано право да не подлежат на чисто автоматизирано вземане на решения.

Понякога работата с различни държави и съответно бизнес култури може да доведе до редица проблеми, например погрешно разбиране на информацията при превод на друг език, субективност, която произтича от някои културни стереотипи, и др. При някои обстоятелства дори показването на различни реклами може да означава, че потребителите на тази услуга се профилират по начин, който се тълкува като дискриминация въз основа на раса, пол, политически убеждения, религия, здравен статус и т.н. Събирането и обработването на данни следва да се извършва по начин, който предотвратява такива дискриминационни последици за лицата и отговаря на принципа на прозрачност на *GDPR*. Затова специалистите по данни трябва да притежават необходимия опит и гъвкавост, за да се справят с различни начини на мислене и правене на бизнес, с оглед постигане на успешен резултат.

Горепосочените стъпки на процеса на работа с големи данни очертават в общ план някои основни трудности, с които се сблъскват специалистите по данни във връзка с изпълнението на принципите на *GDPR* [221]. Управлението на големи данни е трудна задача и за да се постигне успех в тази сложна и силно нелинейна дисциплина, наборът от умения на специалиста по данни трябва да бъде адаптиран към изискванията за защита на данните и моментната ситуация. Етичното поведение в тази нововъзникваща област с толкова много възможности е задължителна част от професионалната характеристика на специалиста, работещ с големи данни, които в повечето случаи имат личен характер.

Съгласно изложените по-горе етични проблеми и предизвикателства, пред които се изправя специалистът по данни, за да постигне съответствие с *GDPR*, ще формулираме необходимите умения, които той трябва да притежава (Таблица 2.7).

Таблица 2.7. Етични умения при работа с лични данни

Задачи в съответствие с <i>GDPR</i>	Дейности на специалистите по данни	Необходими умения
Информирано съгласие	<p>Запознаване с уведомлението, декларациите и процедурите във връзка с поверителното третиране на личните данни.</p> <p>Осигуряване на образец на декларация за съгласие.</p> <p>Получаване на изрично съгласие за всяка дейност по събиране и обработка на данни.</p> <p>Документиране на всички декларации за съгласие и на всички лични данни на потребителите.</p> <p>Събиране на лични данни.</p> <p>Периодичен контрол на валидността на информираното съгласие. Контрол на времето на задържане на данните в зависимост от целта на събирането им.</p> <p>Използване на ясен и изчистен стил при писане, без професионален и технически жаргон.</p>	Бази данни; програмиране; владеене на чужди езици; комуникативни умения
Псевдонимизация и анонимизация	<p>Обработване на лична и чувствителна информация.</p> <p>Използване на подходящи техники за кодиране на събраните лични данни (маски, кодиране, криптиране и др.).</p> <p>Защита на анонимните и псевдонимните данни от неоторизиран достъп и реидентификация.</p>	Програмиране; бази данни; информационна сигурност; мрежи; обработка на данни; вливане на данни (<i>Data ingestion</i>); извличане на данни (<i>Data mining</i>); подготовка на данни (<i>Data preparation</i>)
Профилиране	Оценка за въздействието на <i>Provide data protection impact assessment</i> .	Статистика; обработка на данни; <i>Data ingestion</i> ; <i>Data mining</i> ;

Задачи в съответствие с <i>GDPR</i>	Дейности на специалистите по данни	Необходими умения
	<p>Анализ на данните и откриване на връзки и закономерности.</p> <p>Предоставяне на информация относно профилирането на потребителите.</p> <p>Предоставяне на право на потребителите да не подлежат на чисто автоматизирано вземане на решения.</p> <p>Предотвратяване на ефекти на дискриминация.</p>	<p><i>Data preparation</i>; машинно обучение; комуникационни умения</p>

От Таблица 2.7 става ясно, че етичните умения на специалиста по данни са част от двете категории – **твърди (технически)** и **меки (нетехнически)** умения. Твърдите умения са задължителни за всеки специалист по данни и са свързани с дейностите по събиране, обработка и анализ на големите данни. От друга страна, специалистът по данни трябва да притежава умения за комуникация със субектите на данни, за да се съобрази с техните права съгласно *GDPR*, което е част от групата от меките умения.

През последните години науката за данните все повече навлезе в живота на хората и следователно се нуждае от кодекс за професионална етика, за да установи вида на връзката между професионалистите и обществото (включително клиенти, субекти на изследването, потребители на техните услуги и т.н.). Ясно е, че специалистите, работещи с големи данни, притежават права за достъп и способности за изследване на личните данни на големи групи хора, което повдига нови, неизвестни досега предизвикателства. Във връзка с гореизложеното се налага изводът, че етичните умения са важен елемент от професионалния профил на специалиста по данни.

2.3. ПОТРЕБНОСТ ОТ ЕКСПЕРТИЗА В ОБЛАСТТА НА НАУКАТА ЗА ДАННИ

2.3.1. Търсене на специалисти по данни в световен мащаб

Специалистите по данни са много ценен ресурс за организациите, тъй като са способни да извличат полезно знание от големите данни [105]. Според Wall Street Journal през 2014 г. се наблюдава ръст в търсенето на специалисти по данни [232]. Оказва се, че в този период всички големи компании в света търсят и наемат активно специалисти по данни, за да подпомагат оптималното управление на човешките дейности в най-различни стопански сфери.

Факт е, че организациите използват своите данни, за да подобрят ефективността и ефикасността на операциите си [136]. Осигуряването на човешки капитал с високи аналитични умения за работа с големи данни позволява по-ефикасни и ефективни решения и несъмнено повишава цялостната работа на организациите. Въпреки големия потенциал на новите технологии, инструменти и приложения за анализ, днес основен проблем пред организациите е намирането на служители с необходимите умения [141, 162].

Анализ на IBM показва, че до 2020 г. броят на обявените позиции за специалисти по данни и специалисти по анализ на данни ще достигне 61 799, което представлява едва 2% от прогнозираното търсене за всички работни позиции, изискващи умения за работа с данни и аналитични умения.

Според McKinsey Global Institute [153] броят на необходимите за американската икономика специалисти по данни би могъл да достигне 250 000 до 2024 г. Според доклад на LinkedIn [60] за степента на търсене на новопоявили се свободни работни места в САЩ през 2017 г. специалистите по данни се нареждат на второ място сред 20-те най-търсени професии, веднага след експертите по машинно обучение. Според същото проучване за периода 2012 – 2017 г. има безпрецедентен ръст на търсенето на специалисти по данни и интересът към науката за данните нараства близо 6,5 пъти.

Тенденцията показва задълбочаване на проблема. Търсенето на специалисти по данни не само ще продължи да нараства, но според Davenport and Patil през следващите години се очаква специалистът по данни да бъде най-търсеният професионалист в бранша.

Както става ясно, потребността на организациите от квалифицирани специалисти, притежаващи нужната експертиза за учене от големи данни, не спира да нараства. Основни фактори за увеличено-

то търсене на специалисти по данни са, от една страна, наличието на големи данни, които са твърде обемни и сложни за процесите, използващи традиционните технологии за съхранение и анализ [75], а от друга страна, фактът, че мениджърите разбират, че знанията, получени от натрупаните данни, имат стратегическо значение за развитието на организацията, която представляват.

Организациите трябва да търсят решение в две основни посоки: използване на съвременни технологии и методи и осигуряване на специалисти, които притежават необходимите умения, за да ги използват ефективно и да осигуряват полезно знание [233]. Ролята на специалистите по данни и техните многостранни умения са ключов фактор, който придобива все по-голямо значение в контекста на големите данни. Днес уменията не се разглеждат единствено от гледна точка на експлоатацията на технически ресурси (софтуер и инфраструктура). Резултатът от анализа и обработката на големи данни е свързан с генериране на прогнози, подпомагане на ръководството на организацията във вземането на информирани решения и осигуряване на конкурентоспособност на организацията [202].

За получаването на смислени и полезни резултати при обработката на данни, а също и на рационални изводи при наличие на несигурност, се изискват разнообразни умения. Експертите трябва да притежават широк обхват на компетенции, както стана ясно в предходната точка. Освен това, поради динамичното развитие на науката за данните, е необходимо те постоянно да бъдат развивани, разширявани и надграждани. Понякога съществуват и допълнителни изисквания към уменията на специалистите по различни езици за програмиране и технологии, които са специфични за организацията.

Съвременни изследвания [126] показват, че специалистите по данни са високообразовани – 88% имат поне магистърска степен, а 46% имат докторска степен. Макар да се срещат изключения, обикновено високата степен на образование е необходима, за да се развие дълбочината на знанията, нужна за постигането на уменията на учен – експерт в областта. Най-често специалистите по данни притежават бакалавърска степен по специалностите „Компютърни науки“ или „Статистика“.

Разпределението на специалистите според завършената от тях специалност показва, че най-голям дял заемат специалистите, завършили специалностите „Математика“ и „Статистика“ (32%), следвани от тези, завършили „Компютърни науки“ (19%) и „Инженеринг“ (16%).

В следващите точки ще разгледаме търсенето на специалисти по данни на българския пазар и изискванията, които организацията поставят към техните знания, умения и опит. Основната цел на това изс-

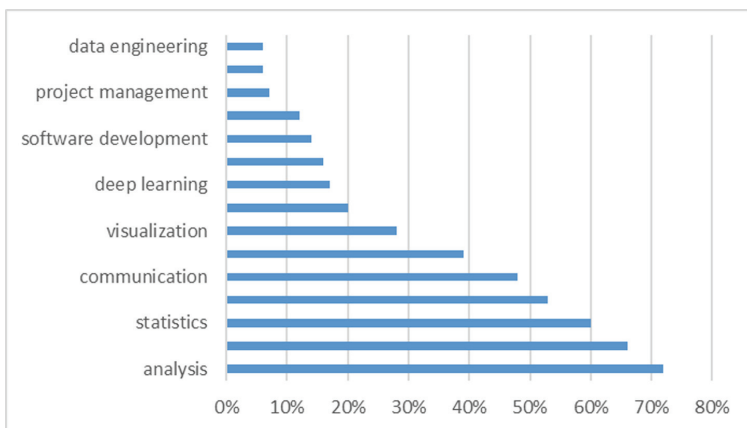
ледване е да съпоставим създадената рамка на професионалния профил на специалиста по данни, която се основава на анализ на литературата, с професионалния профил на специалиста, който търсят организациите на пазара на труда.

2.3.1. Търсене на специалисти по данни в България

А. Изисквани умения

С оглед проучване на търсените умения от страна на бизнес организациите беше извършен преглед на активните обяви за работа в най-популярните онлайн сайтове за работа в България, в това число на знания, умения и опит, които трябва да притежават специалистите, изисквания за владене на програмни езици и инструменти. Близко 100% от обявите за търсене на „специалист по данни“ са публикувани от международни компании, функциониращи на българския пазар.

Прегледът на обявите за работа показва, че търсените от бизнеса умения може да се групират в общи категории, включващи анализ, машинно обучение, статистика, компютърни науки, комуникация, математика, изкуствен интелект и др. (Фиг. 2.7).



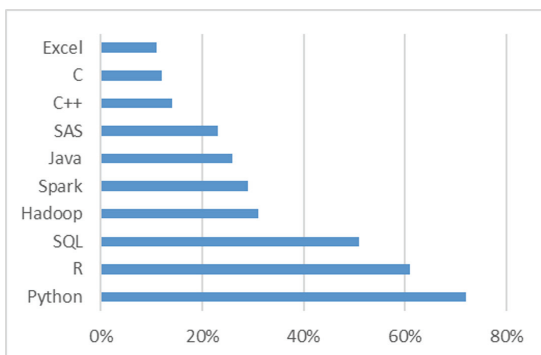
Фиг. 2.7. Задължителни умения в професионалния профил на специалиста по данни според работодателите

Анализът и машинното обучение са основни умения, изграждащи профила на специалиста по данни. Извличането на закономерности и модели от данни и прогнозирането са основна функция на науката за данните, а машинното обучение включва умения за осъществяване на прогнози.

Аналитичните умения са важен фактор, особено при работа с големи данни. В повече от половината от обявите за работа се акцентира върху комуникационните умения и уменията за работа в екип.

Въпреки че не се срещат толкова често в обявите за работа, изкуственият интелект (*AI*) и *Deep Learning* може да се разглеждат като част от машинното обучение. Това се дължи на факта, че *Deep Learning* се използва все по-често при машинното обучение и при обработката на естествени езици.

В по-голямата част от обявите за работа работодателите изискват умения за работа с конкретни софтуерни инструменти, както и знания за работа с различни програмни езици, библиотеки и технически инструменти (фиг. 2.8).



Фиг. 2.8. Търсени от бизнеса технически умения на специалиста по данни

Най-търсеният език за програмиране в обявите за работа е *Python*. Популярността на този език с отворен код продължава да расте. Една от причините е, че той е приятелски настроен за начинаещите програмисти, предлага много ресурси и помощ при самообучение. Освен това голяма част от инструментите, прилагани в науката за данните, са съвместими с него. Следователно може да се твърди, че днес *Python* е език за програмиране, който е задължителна част от професионалния профил на специалиста по данни.

Както се вижда на Фиг. 2.8, езикът *R* не изостава много от *Python*. Доскоро се смяташе, че *R* е основният език за специалистите по данни, но постепенно *Python* започна да го измества. Силата на този език за програмиране с отворен код е при работата със статистически данни, поради което все още е много популярен сред статистиците.

В списъка със специфични изисквания на работодателите фигурира и *SQL*. Като структуриран език за заявки, той е основният начин

за взаимодействие с релационни бази данни. Напоследък като че ли се забелязва намаляване на търсенето от работодателите на *SQL*, но въпреки това уменията за работа с този език не са за пренебрегване.

Казаното по-горе потвърждава изследване на Valchanov, според което представителят на професията „специалист по данни“ [225]:

- най-често е от мъжки пол – оказва се, че професии, свързани с науката за данните, са силно доминирани от мъже, които представляват 70% от извадката на изследването;
- владее поне един чужд език;
- притежава магистърска или докторска степен;
- има стаж, не по-малко от 4,5 години след придобиване на образователната степен;
- предпочита програмните езици *R*, *Python* и *SQL*.

Въз основа на горезложеното може да се направи изводът, че професионалният профил на съвременния специалист по данни включва следните задължителни умения и компетенции:

- *аналитично и критично мислене* – специалистите по данни трябва да притежават аналитично, логическо и критично мислене, за да извършват обективен анализ на фактите по дадена задача или проблем, преди да правят преценки и да формулират становища;
- *програмиране* – техническите умения по програмиране са задължителни за специалистите от този вид;
- *статистика и математика* – задълбочените математически знания, както и статистическите, се използват при разработката на модели;
- *комуникационни умения* – специалистът по данни трябва да умее да си сътрудничи тясно с ръководителите на организациите. Също така е необходимо да умее да обяснява сложните процеси в „черната кутия“ по прост и разбираем начин, да дава прогнози, препоръки и съвети, за да подпомогне ръководството в процесите на вземане на решения.

Освен тези умения, които изискват преобладаващата част от организациите, в обявите за работа се наблюдава разграничаване на базата на опитност и експертиза. Преглед на тази диференциация ще бъде извършен в следващите редове.

Б. Изискван опит

Анализът на обявите за работа показва, че фирмите се нуждаят от специалисти по данни с различна степен на експертиза и опитност. Работодателите правят разлика между „младши учен по данни“ (Junior), „специалист по данни от средно ниво“ (Mid-Level) и „старши специа-

лист по данни“ (Senior). В Таблица 2.8 е направено сравнение на изискванията за опит и образование между трите категории специалисти.

Таблица 2.8. Нива на експертност в областта на науката за данните

Категория	Опит (в години)	Образователна степен
Младши специалист по данни	От 0 до 2 години	Минимум бакалавърска степен в областта на статистиката, математиката или компютърните науки
Специалист по данни от средно ниво	3 и повече години	Магистърска или докторска степен
Старши специалист по данни	5 и повече години изследователски опит	Докторска степен в областта на статистиката; сертификати за умения

Основното различие между трите равнища в кариерното развитие на специалиста по данни е свързано, от една страна, с възможностите за самостоятелност и независимост при изпълнение на поставените задачи. От друга страна, различието е провокирано от възможностите за организиране на работния процес в съответствие с нуждите и сроковете, поставени от компанията.

По-съществени се оказват различията между младшия специалист и специалиста от средно равнище, поради което в настоящата секция ще бъде направено сравнение между тези две нива на експертност.

Най-общо казано, старши специалистът по данни добре разбира, че данните винаги съдържат недостатъци, породени в процеса на създаването и начина на записването им, и разбира се, от субективната преценка на осъществяващите тези дейности. Тези недостатъци влияят на резултатите от изследването, получени при анализа на данните. Освен това старши специалистът по данни притежава меки и твърди умения. Този вид специалисти отговарят за организирането и автоматизирането на задачите и за възлагането на задачите на младшите членове на екипа или на външни консултанти. Част от многото други отговорности, пред които е изправен старши специалистът по данни, включва [160, 161]:

- ръководство на хора, подбор и назначаване на подходящи хора, ръководство на ръководителите, които отговарят пред него;
- обучение на колегите, които не са технически сръчни;
- избор на правилните инструменти и преценка на ползите и минусите на предлагания софтуер и платформи за специфичните мащабни проекти;

- определяне на правилните алгоритми и статистически техники с отчитане на спецификата на дадения проект;
- определяне на полезните външни и вътрешни източници на данни, комбиниране на различните източници при изчистване на данните;
- разбиране на жаргона на изпълнителните мениджъри и транслиране на техните изисквания, съображения или идеи в успешни приложения;
- оценяване на възвръщаемостта на инвестициите за компанията;
- успешно комуникиране с ръководителите/колегите/ мениджърите/клиентите от всеки тип, което включва основни комуникативни и поведенчески умения.

Опитът и експертизата във всяка професия се изграждат в процеса на нейното практикуване. Профилът на специалиста с натрупана експертиза не е привлекателен при изграждането на учебен план за обучение в магистърска степен в областта на науката за данни. За нас по-интересни са специалистите, попадащи в графата „младши специалист“. Това са специалистите без стаж и опит, които завършват своята бакалавърска или магистърска степен на образование и търсят възможности да се реализират по професията.

V. Обобщен професионален профил

В следващата таблица са представени обобщени данни от напредното изследване, което ще ни послужи като основа за изграждане на рамка на търсения профил на специалиста по данни. В Таблица 2.9 са разгледани посочените от работодателите изисквания по отношение на следните показатели: степен на образование, твърди (технически) умения, меки умения, аналитични умения, езици и опит.

Таблица 2.9. Знания и опит, изисквани от работодателите

Образователна степен	Технически умения	Меки умения	Аналитични умения	Опит
Бакалавърска степен в областта на статистиката, приложената математика, компютърните науки или други свързани области	Скриптов езици (Python, R) Езици за обектно ориентирано програмиране (Scala, Python, Java, C++)	<i>Комуникационни умения</i> – способността да представя сложната техническа информация на прост, лишен от терминология език	Творческо мислене Умения за анализиране и решаване на проблеми	От 2 до 4 години

Образователна степен	Технически умения	Меки умения	Аналитични умения	Опит
Магистърска степен в областта на статистиката, компютърните науки и свързани с тях области	Бази данни <i>Big Data frameworks</i> (<i>Hive, Spark, Hadoop</i>)	<i>Организационни и лидерски умения</i> – самотивиращ се и организиран		
Докторска степен в областта на статистиката, математиката, компютърните науки или свързани с тях области	Статистика (<i>SPSS, SAS</i>) <i>Техники за машинно обучение</i> (кълстерен анализ, дървета на решенията, невронни мрежи и др.) <i>Статистически техники и концепции</i> (регресионен анализ, статистически тестове и др.)			

Отчитайки фактора образование и квалификация, организациите обикновено изискват кандидатът за позицията „специалист по данни“ да бъде най-малко с бакалавърска степен (около 24% от обявите за работа). Най-често поставяното условие по същия показател се оказва притежаването на магистърска степен (62%), а най-рядко посочваното е специалистът по данни да притежава докторска степен (едва 14%). Това, което обединява всички обяви по показателя „образователна степен“, е специалността: всички, търсещи такива кадри, изискват кандидатът да бъде специалист в областта на статистиката, математиката, компютърните науки и свързаните с тях области.

По отношение на фактора технически умения търсенето ограничава обхвата на знанията на специалиста по данни най-често до скриптов езици, езици за обектно ориентирано програмиране, фреймуърци на големи данни, статистика, техники за машинно обучение, статистически методи и концепции.

В своите обяви работодателите държат най-често кандидатът да притежава меки умения, включващи комуникационни, организационни и лидерски умения.

След като получат изброените в Таблица 2.9 квалификации, кандидатите трябва да са придобили опит между 2 и 4 години.

Според изследване на Kdnuggets популярността на някои софтуерни инструменти нараства пропорционално, докато при други има рязък спад [182]. Картината, представена в това изследване, може да ни подсказва какви ще бъдат тенденциите през идните няколко години и съответно какви инструменти трябва да бъдат заложени в учебните курсове, подготвящи специалисти по данни. Изследването потвърждава, че най-търсените през последните няколко години технически умения се свързват с езиците *Python*, *R* и *SQL*, които на този етап участват в предложения списък с технически умения.

В наши дни съществува ново дигитално разделение между тези, които отговарят на показателите, изисквани от бизнес сектора, и тези, които не отговарят на пълния набор от изисквания. Въпреки че много университети и академии предлагат курсове и магистърски програми за обучение в областта на науката за данните, днес на пазара на труда има недостиг на тези специалисти. Освен това на фона на голямото търсене на специалисти с необходимия опит в областта на науката за данните и въпреки производството на нови кадри от университетите, днес съществува пропаст между търсенето и предлагането. Оказва се, че малко специалисти напълно отговарят на нуждите на бизнеса. В резултат на това се наблюдава дигитално разделение на ниво умения, знания, квалификация и опит за работа с големи данни [194].

РЕЗУЛТАТИ И ИЗВОДИ ПО ГЛАВА II

В тази глава разгледахме професионалния профил на специалиста по данни от гледна точка на знанията, уменията и компетенциите, които той трябва да притежава, за да обработва и анализира големи данни. Във връзка с поставените задачи може да се посочат следните резултати от глава II:

1. Направен е преглед на различните гледни точки за компетентностите, които специалистът по данни трябва да притежава.
2. Въз основа на литературен преглед е изведен обобщаващ професионален профил на специалиста по данни.
3. Резултатите са анализирани и обобщени.
4. Определени са областите от знания и умения, които са съществени за успех в професията.

ГЛАВА III.

ОБУЧЕНИЕ В ОБЛАСТТА НА НАУКАТА ЗА ДАННИТЕ

Съвременните технологии, инструменти и приложения за обработка и анализ на големи данни предлагат добри възможности за извличане на полезни знания. Проблемът, пред който днес са изправени различните организации, е свързан с намирането на професионалисти с необходимите умения, които да реализират тези възможности. Очаква се търсенето на специалисти в областта на науката за данните да продължи да нараства през следващите години паралелно с нарастването на разбирането за възможните ползи, които може да се извлекат от натрупаните данни.

На практика, на голяма част от обучаващите организации се налага да догонват темпа на развитие на технологиите и динамиката в областта на науката за данните, за да предложат обучение, което да отговаря на пазарните нужди. Постигането на компетентност за учене и работа с големи данни е задача, която изисква реинженеринг на съществуващия начин на планиране, изграждане на учебни програми и осъществяване на обучение.

В тази глава се разглеждат основни аспекти, свързани с обучението на студенти по програми в областта на науката за данните.

3.1. ПРОЦЕС НА ОБУЧЕНИЕ В ОБРАЗОВАТЕЛНИТЕ ИНСТИТУЦИИ

3.1.1. Таксономии за целите на образованието

Предмет на педагогическите таксономии са определянето и систематизирането на целите на обучението, реализацията на които води до мислене от високо ниво. Според [15, 196] мисленето от по-високо ниво се характеризира с: липса на алгоритъм (пътят на дейността не е предварително напълно определен); комплексност (пътят не е видим от една-единствена гледна точка); наличие на множество конкурентни решения (всяко с различна стойност и предимства); нюансирана преценка и тълкуване; прилагане на множество критерии (някои в конфликт помежду си); безпорядъчна структура; несигурност и саморегулиране на процеса. Когато човек има умения за мислене от високо

ниво, той не само разбира същността на изучаваните предмети и явления, но и открива закони и закономерности, принципи и правила, разкрива тенденции за развитие и бъдещо поведение, участва в конструктивна, съзидателна и творческа дейност, съзнателно и разумно ръководи и контролира собственото си индивидуално и социално поведение. Таксономите играят съществена роля в дидактиката и педагогическите науки, тъй като подпомагат процесите на [15]:

- правилно определяне и степенуване на целите и задачите на обучението;
- разностранно планиране на дейностите в процеса на обучението;
- обособяване на познавателните структури в обучението;
- формулиране на проблеми и поставяне на задачи пред обучаемите;
- подобряване на адекватните на поставените цели инструменти за оценка;
- коригиране на процеса на обучение на базата на резултатите от обучението и изпитанията от обучаемите трудности

През 1956 г. Benjamin Bloom за първи път представя теория за таксономия, свързана с целите на образователната система и обучението [41]. Той е смятан за „баща“ на идеята за формиране на интелектуални умения в обучението. В днешно време таксономията на Bloom се разглежда като таксономия за умения за обработка на информация и често се използва при проектирането на образователни и учебни процеси. Нивата в таксономията са подредени от лесно към сложно и от конкретно към абстрактно, като всяко ниво трябва да бъде овладяно, преди да се премине към следващото, по-високо ниво (Фиг. 3.1).

Според таксономията на Bloom мисленето може да се класифицира в шест когнитивни нива на сложност: *знания, разбиране, приложение, анализ, синтез и оценка*, както следва [41]:

- **Знания:** Включва запомняне, възпроизвеждане и разпознаване на факти, термини, закони и дефиниции в същата форма, в която са преподадени в учебната среда. Това е равнището на догматично усвояване на информацията и от обучавания не се очаква да интерпретира или преработва знанията, а само да ги запомни като задължителни факти. Това най-ниско ниво на познавателната дейност е основополагащо за всички следващи категории и без него обучението би било невъзможно.
- **Разбиране:** Включва способността за извличане на същест-

вен смисъл от изучаваната материя. Това е трансформирането на възприетите знания чрез обяснение, обобщение и прогнозиране. Нивото на разбирането използва информацията, натрупана като резултат от възприемането.

- **Приложение:** Включва способността за използване на „разбраното“ в нови, непознати ситуации. На това ниво вече „разбрани знания“, като принципи, методи, правила и концепции, се прилагат за решаването на непознати задачи.

- **Анализ:** Включва разделяне на общата материя на по-малки части и разглеждането им поотделно като самостоятелни единици. Това ниво е стъпката към решаването на по-сложни задачи.

- **Синтез:** Включва способността за обединение на вече анализирани части в едно цяло. При нивото на синтез вече може да говорим за „творческо мислене“, понеже частите се обединяват по съвсем нов, непознат до този момент начин.

- **Оценка:** Включва способността за обобщаване и окачествяване на стойността на решена (синтезирана) задача. Оценката винаги трябва да се дава по точно определени критерии и норми.

Таксономията на Bloom представлява йерархия на мисловните умения, в която по-високите нива на мислене включват всички познавателни умения от по-долните нива, затова се представят във вид на йерархия, най-често пирамида (фиг.3.1а). Тя намира широко приложение в образователната сфера в много държави, както и в България.



a) Таксономия на Bloom



b) Модифицирана таксономия на Anderson и Krathwohl

Фиг. 3.1. Йерархия на мисловните умения

Таксономията на Anderson и Krathwhol (2001), показана на фиг. 3.1b), е позната като „модифицирана таксономия на Bloom“ [25]. В тази таксономия броят на нивата отново е шест, като най-ниското ниво е преименувано от „знания“ в „запаметяване“, а наименованията на останалите също са променени от съществителни в глаголи, за да се постави акцент върху действието в процеса на мислене. По този начин се прави ясна разлика между съдържанието на мисленето и процедурите за решаване на проблеми. Освен това са разменени местата на последните две нива, като синтезът (наречен „създаване“) е поставен на по-високо ниво от оценяването. С други думи, творческите способности и креативността представляват най-високо ниво в процеса на обучение, тъй като те са свързани с формулирането на нови, оригинални идеи, разкриването на важни взаимовръзки между предмети, явления, откриването на нови начини за решаване на задачи и проблеми, възникващи в теорията и практиката. Във връзка с динамично развиващите се цифрови технологии и в контекста на големите данни днес креативността е много важно и търсено качество, нов тип актив за организациите, който повишава тяхната конкурентноспособност.

Таксономията на Bloom се свързва и със съответни ключови думи - глаголи, които са специфични и характеризират всяко ниво. Тези специфични глаголи се използват в дейностите, които се задават при в съответно нивот В Таблица 3.1 са представени основните компоненти от таксономията на Bloom в посока от репродуктивност към продуктивност.

Таблица 3.1. Таксономия на Bloom

Ниво	Глаголи
Запаметяване	Назовава, описва, проследява, разграничава, изброява, съвпада, Кой, Какво
Разбира	Разпознава, обяснява, илюстрира, дискутира, интерпретира, класифицира, сравнява
Прилага	Планира, характеризира, експериментира, изчислява, моделира, прилага, избира, построява, разработва
Анализа	Изследва, доказва, прави хипотези, сравнява, прави заключения
Оценява	Оценява, аргументира, обобщава, награждава, избира, създава критерии, критикува
Създава	Създава, конструира, проектира, адаптира, комбинира, развива, формулира, изгражда

Таксономията на Guilford (1955) се базира на описателно-обяснителната представа за структурата и функционирането на интелекта. В нея се разграничават три основни структурни звена:

- **Операции:** Включва основни интелектуални процеси, използвани от човека.

- **Резултати:** Включва съдържанието на интелекта, т.е. „това, което мислим“, във вид на единици, класове, отношения, системи, преобразувания и следствия.

- **Способи:** Включва начини за представяне на резултатите като образ и съдържание.

Guilford класифицира цели 150 фактора в съдържанието на тези три структурни звена. Всеки от тях се постулира като независим, но с различна степен на независимост, което е и слабата страна в тази теория. Таксономията на Guilford не изключва възможността да бъде използвана в синхрон с таксономията на Bloom.

Съществуват и други подобни таксономии, например на Dave (1970), Morzano (2000), Fink (2003) и др. Всички гореизброени таксономии се използват масово за подпомагане на педагогическата наука и реализиране на ефективни образователните процеси.

3.1.2. Предизвикателства пред обучаващите институции

Днес развитието на всички стопански сфери силно зависи от данните и компетенциите за тяхната обработка и анализ. Данните са навсякъде около нас, а обемът им нараства експоненциално, затова се налага те да се доставят, обработват, извличат и управляват, да се общува посредством тях, да се вземат решения въз основа на тях и т.н. Все повече комерсиалните и публичните организации осъзнават стойността на големите данни и възможностите да се извличат полезно знание и добавена стойност от тях.

Напоследък често се срещат концепции за основани на данни (*Data Driven*) училища, организации и др. Културата, основана на данни (*Data Driven Culture*), се превръща в ключов фактор за организациите, които днес имат възможност да използват предимствата на цифровото пространство и на големите данни. Booth и Hendrix изследват процесите, свързани със създаването на *Data Driven Culture*, посочвайки седем ключови области, към които трябва да се насочат усилията:

- **култура** – когато културата, основана на данните, не доминира, често се поражда недоверие към качеството на данните и начина, по който отразяват реалността, към техниките за анализ и начините

на представяне;

- **таалант** – липсват подходящи специалисти, боравещи с данни, тъй като уменията и знанията, които те трябва да притежават, са многостранни и рядко се съчетават при един човек;

- **разходи** – технологиите и инструментите за работа с данни изискват влагане на средства, които обаче може да се разглеждат по-скоро като стратегическа инвестиция;

- **собственост на данните** – липсата на централизирано съхранение на данните е пречка, тъй като данните се съхраняват в индивидуални силози;

- **качество на данните** – за успешното реализиране на процесите на извличане на полезно знание от данните е необходимо да се гарантира точност на данните;

- **стандартизация на данните** – организациите трябва да изграждат речници, за да преодоляват рисковете от грешно разбиране и интерпретиране на данните, и да направят процеса на анализа им по-разбираем;

- **достъп до данните** – препоръчва се данните да се съхраняват централизирано, а индивидуалните силози да се премахнат.

Непрекъснатият прогрес в областта на ИКТ открива възможности за съхраняване, обработка, предоставяне на достъп и извличане на полезно знание от много големи информационни масиви. Масовото навлизане на технологиите в живота и икономиката трансформира възможностите за работа, обучение, общуване, достъп до информация и прекарване на свободното време. В резултат от това е налице глобална електронна среда, която предоставя нови възможности за комуникация и взаимодействие на индивидите и общностите в световен мащаб.

През последните години се промени основният принцип на изграждане на компютърно базирани информационни системи, които организират информационните ресурси и обслужват потребителите. Наблюдава се еволюция на информационните системи – от системи, доставящи данни, до системи за ефективно информиране, които подпомагат потребителите и организациите в решаването на различни задачи. С цел намаляване на времето за обучение на потребителя, още в процеса на проектиране на информационната система се обръща внимание на значимостта на „удобството“ на потребителя при работа с нея. Но въпреки това грамотността и компетентностите на потребителите са от решаващо значение за успеха при работа с данни и информация.

Вследствие на гореизложените фактори и тяхното постоянно развитие и предефиниране, пред образователната система се разкриват много нови възможности, но и много нови проблеми и предизвикателства. Тя е изправена пред нелеката задача да осигури на организациите кадри, които да отговарят на техните изисквания за умения, компетенции и опит. Пораждат се въпроси като следните:

(1) От специалисти с какви умения се нуждае бизнес секторът?

(2) Отговарят ли днес университетските програми и практики на очакванията на бизнеса и подготвят ли добре бъдещите специалисти по данни?

(3) Съществува ли разминаване между търсенето (от индустрията) и предлагането (от университетите и други обучаващи организации) на специалисти в областта на науката за данните?

В ерата на големите данни се създават предпоставки за осигуряване на качествено и модерно образование, осигуряват се възможности за реализация и повишаване на конкурентоспособността на работници и работодатели на свободния пазар. Освен на техническите аспекти, свързани със съхранението и извличането на данни, днес се акцентира и на начините за използването на получените данни. Във връзка с това образователните институции се изправят пред следните проблеми:

- липсата на зрялост и бизнес опит у обучаваните пречи на способността им да оценяват нивото на полезност и приложимост на получените данни;

- липсата на информация, статистически и математически знания и умения у обучаваните пречи на способността им да преценяват дали дадена техника е приложима и дали резултатите, получени чрез проучване на дадена техника, са валидни в контекста на областта на данните.

И двата посочени проблема представляват значителна бариера при обучението в компетенции, свързани с обработка и анализ на големи данни. Освен това образователните институции трябва да се развиват своевременно, като съчетават традиционните подходи и дисциплини с бързия темп на развитие на ИКТ, в частност разработването на нови приложения и инструменти за работа с големи данни, нови подходи към базите данни (нерелационни бази данни, разпределени бази данни, облачни технологии и др.) и т.н. Образователните институции, естествено консервативни заради дългия период, необходим за потвърждаване на правилността на взетите решения, са изправени пред темп на развитие, на който трудно могат да отговорят.

Според Chen et al., 2012 от създаването си преди около 50 години „Информационни системи“ като академична дисциплина се фокусира предимно върху нуждите на бизнеса, които включваха основно управление на вътрешните данни и трансакции за организациите [50]. Днес възниква необходимост от нова визия за преподаването на „Информационни системи“ и въвеждането на дисциплини за обучение по извличане на данни, текст и мнения, анализ на социални медии и мрежи, уебмайнинг и прогнозен анализ, които са необходими за изграждането на специалисти по данни, подпомагащи организациите в процесите на вземане на решения. В университетите се осъзнава нуждата от обучение в областта на обработката и анализа на данни, дори и при изучаването на нетехнически специалности. През 2015 г. в учебния план за общо образование на Temple University, САЩ, е въведен избираем курс по наука за данните, който е адресиран към студенти от разнообразни специалности. Този курс има за цел изграждането на грамотност за работа с данни, която е свързана с традиционната представа за мисията на компонента „общо образование“, типичен за американския образователен модел:

1. Информационна грамотност, включително способност за разпознаване и формулиране на информационни потребности; намиране, критично оценяване и организиране на информация за конкретна цел; етично използване на информацията.

2. Развитие на умения за критично мислене, включително доказателства, оценка, анализ и синтез на множество източници и разсъждения върху различни перспективи.

3. Комуникационни умения – използване на говорим и писмен език за конструиране на съобщение, което демонстрира ясни цели за целевата аудитория.

4. Извличане, организиране и анализ на данни, свързани с научен модел.

5. Разбиране и комуникиране на това как информационните технологии подпомагат процеса на откриване на иновации.

6. Използване на научно или технологично мислене за решаване на проблеми, които са част от ежедневието.

Според Dichev [81] подобен курс се предлага и в друг университет, Southern Liberal Arts University, САЩ, който предлага типично *Liberal-arts* образование. Курсът е насочен към изграждане на грамотност за работа с данни на нетехническа аудитория с акцент върху събирането, обработката, анализа и използването на данните за информирано и етично вземане на решения. Целта му е да постави науката

за данните в основата на общото образование, да подготви студентите в начален етап от обучението им за промените в обществото и технологиите, управлявани от данни, и да подпомогне ролята на висшето образование в изграждането на критично мислене, информационна и количествена грамотност. Грамотността за работа с данни съгласно този курс включва:

- способност за формулиране на правилни въпроси;
- способност за количествено мислене;
- способност за аналитично мислене;
- способност за визуализиране и докладване на данните.

Основните знания, които се преподават с цел развиване на тези умения, са:

- изчислителни методи;
- събиране и обработка на данни;
- моделиране на данни и статистически анализ;
- визуална комуникация и визуализация.

Като обединяваща рамка за интегриране на тези теми е избран езикът за програмиране *Python*. Курсът е адресиран към аудитория, която не е с техническа насоченост. Затова се различава от типичните въвеждащи курсове по програмиране – въвеждат се общи представи за програмирането с акцент върху подготовката, обработката и представянето на данни.

В заключение може да се каже, че анализът на данни вече не е привилегия само на специалистите по данни. Това е основно умение за 21. век, което е едновременно предизвикателство и възможност за висшето образование. Умението трябва да се развива у студентите с цел извличане на полезно знание и прозрения.

3.1.3. Обучение, базирано на компетенции

В свое изследване Окуловский [11] разглежда същността на понятието „компетентност“ и връзката му с резултатите от обучението. Авторът разглежда мястото на личните качества, притежавани от специалиста по данни при решаването на конкретен казус. Според него „компетенциите могат да бъдат представени като способността за извършване на конкретни дейности в конкретна област, основаваща се на прилагането на знания и умения при едновременното прилагане на лични качества, които правят тази дейност успешна“. Същевременно Окуловский представя компетентността като „способност (и желание) да бъдат извършвани професионални дейности в определени облас-

ти въз основа на прилагане на придобитите от обучението компетенции“. Съвременният подход в обучението изисква резултатът от обучението да не е просто овладяването на конкретни знания и умения, а да бъде фокусиран върху изграждането на способности за тяхното прилагане. Все по-често се говори за функционално-действен подход в образованието, който Окуловский определя като най-важния принцип на съвременното обучение.

Връщайки се към понятията „компетенции“ и „компетентност“, трябва да отбележим, че компетентностите са резултат от пряко и косвено въздействие на множество фактори (от образователен характер) върху обучаемите. Татур отбелязва, че: „...*нито отделните академични дисциплини, нито дори съдържанието на учебния план могат да бъдат „отговорни“ за формирането на определени компетенции*“ [14]. Компетенциите зависят от образователните технологии, методи, организационни форми, образователна среда и др.

Компетентността на индивида е потенциална и относителна. За реалното ниво на нейното формиране може да се съди само тогава, когато завършващият преминава към своята професионална реализация. На практика в процеса на университетското обучение може да се оценяват само придобитите компетенции.

В педагогическата наука възприетите компетенции се разглеждат като основни компоненти на потенциалната компетентност, която завършващият университет студент може да демонстрира в бъдеще [9, 10, 13, 14]. Следователно моделът на компетентност на завършващ своето образование студент не е равен на модела на „специалист“, тъй като компетентността е неразривно свързана с опит, който не може да бъде придобит в нужното количество по време на обучението в университет.

Подходът, основан на компетенциите, е подход, който се фокусира върху резултата от обучението. Резултатът не се изразява в количеството научена от студента информация, а в неговата способност да действа адекватно в различни ситуации (включително в ситуации на несигурност). С други думи, обучението, базирано на компетенции, се фокусира върху това, което учащият ще може да направи или приложи в професионален контекст. Подходът, основан на компетенции, предполага наличието на система за формиране на професионална компетентност у завършилите.

За да може да решава професионални проблеми, специалистът трябва да притежава компетенции, съответстващи на конкретни професионални задачи (основните компоненти на които са знанията,

уменията и личните му качества). Уменията, които позволяват покриването на пълния списък от професионални задачи в определена област, представляват компетентност на специалист в тази област. Ето защо при прилагането на подхода, основан на компетенции, основният акцент в образователния процес трябва да е върху формирането на професионална компетентност на студентите в целия спектър от професионални задачи, изисквани от специалист в неговата професионална дейност.

Основните стъпки в процеса на разработване на обучение, базирано на компетенции, са:

- разработване на компетенциите;
- съгласуване на компетенциите;
- определяне на участниците в изследването;
- изготвяне на информационен материал и информиране на участниците;
- избор и адаптиране на методологията и инструментариума;
- оценяване на компетенциите;
- обобщаване на резултатите;
- изготвяне на програми за развитие и обучение.

3.1.4. Разработване на компетенциите

При изграждането на обучение, базирано на компетенции, се очаква да бъде осъществен преглед както на основните, така и на специфичните компетенции, характерни за конкретни позиции. В таблица 3.2 е представен списък на общите компетенции, участващи в профила на специалиста по данни. Този списък може да бъде приложен върху широк спектър от специалисти в различни сфери на дейност.

Таблица 3.2. Общи компетенции [16]

Наименование	Дефиниране на дейността
Планиране и организиране	Подготвя ясни краткосрочни и дългосрочни планове; ефективно подрежда приоритетите; наблюдава напредъка и осигурява изпълнението на следващите стъпки; прилага систематичен подход.
Практическо решаване на проблеми	Анализира и разбира проблема (ситуацията), като го разделя на по-малки, обозрими за обхващане елементи и части, или проследява стъпка по стъпка произтеклите последици; организира и свързва методично елементи от проблема/ситуацията; проверява причините; систе-

Наименование	Дефиниране на дейността
	матично сравнява различни характеристики и аспекти; организира приоритетите на базата на аргументи, разглежда събитията, разположени хронологично във времето, както и причинно-следствени връзки или отношения („ако..., то...“).
Изготвяне на оптимизирани решения	Поставя фокус върху спазването на закона, процедурите и правилата при подготовката на решения; ориентира решенията към максимален резултат с минимални разходи на ресурси.
Анализиране на данни и информация	Анализира данни и информация и прави изводи за очертаващи се тенденции; предлага и взема аргументирани решения, произтичащи логично от направения анализ.
Аналитично мислене	Подбира подходящите източници на информация; оценява критично и анализира точно и логично комплексни данни; анализира цялата събрана информация, когато взема решения, и дава препоръки.
Умения за междуличностна устна комуникация/общуване	Осигурява информация, от която останалите да могат да разберат ситуацията; умее да изслушва и разбира позициите на различните страни и да предприема своевременно и ефективно необходимите действия; може да приема убедено и доброволно преосмислянето и промяната на собствената си позиция с цел подобряване на крайния резултат; разбира и предоставя информация на други хора на различни нива в организацията и извън нея (служители и клиенти) и желае да се учи от действията на другите и от своите собствени грешки.
Методичност при изпълнение на задачите	Организира задачите и времето си ефективно; изпълнява рутинната си работа съвместно и последователно; обръща внимание на подробностите и следи за спазването на процедурите.
Работа в екип и сътрудничество	Работи в сътрудничество с други колеги, а не самостоятелно или в конкуренция с тях; съобразява се със и прилага правилата на екипа; насърчава и подпомага останалите и участва в открит, непрекъснат обмен на информация и мнения, различаващи се от останалите; екипът може да е в рамките на един отдел, дирекция или организация.

За изследване на специфичните компетенции, които трябва да притежава специалистът по данни, ще бъдат използвани моделите на Националната система за оценка на компетенциите (<https://mycom>)

retence.bg), която функционира успешно на национално равнище. Системата предоставя яснота за необходимите компетенции за работа в сектора в дадена професионална роля и описва нужните компетенции за успешно справяне на работното място.

Информационната система за оценка на компетенциите (ИСОК) представя достъп до модели, състоящи се от основни компоненти. Моделът на длъжност в ИСОК включва:

- **описание на длъжността** – цел, задачи, условия на труд, минимален осигурителен доход, необходими знания и умения, изисквано квалификационно ниво, професионален опит и допълнително обучение;

- **компетентностна рамка**, съдържаща от 9 до 12 компетенции, които са разписани с дефиниция, тип, нива на проявление и поведенчески индикатори. Последните са образци на поведения (в контекста на сектора), извлечени непосредствено от практиката, чрез които се измерва нивото на притежание на компетенцията. Поведенческите индикатори илюстрират как компетенцията от абстрактно наименование се превръща в конкретни, ясни и измерими действия, описващи профила на успешния служител (работник), заемащ длъжността. Компетенциите са класифицирани в три групи – общи за сектора, специфични за длъжността и управленски (ако позицията изисква такива).

Това, което трябва да се уточни, е, че моделът на длъжност не е длъжностна характеристика. *„Длъжностната характеристика определя КАКВО трябва да върши служителят (работникът), докато компетентностният модел на длъжността подсказва КАК трябва да го върши, какви поведения са приемливи за работодателите в сектора и с какво се отличава успешното от посредственото представяне“.*

Информационната система за оценка на компетенциите позволява търсене на длъжност чрез избор на една от възможностите: сектор, код на професията по Националната класификация на професиите и длъжностите (НКПД), както и по ключова дума. Търсенето в системата показва наличието на две длъжности, чиито задължения включват анализ на данни:

- **Бизнес анализатор, информационни технологии** [Информационни технологии] [НКПД код: 25116002] – „Подпомага разработката на нови или модификацията на съществуващи ИТ системи за различни клиенти, като анализира бизнес потребностите им в детайл. Документира данните за идентифицираните потребности и направените анализи и подпомага екипите в компанията при разработването на крайните решения“.

- **Експерт, анализ и дизайн** [Информационни технологии] [НКПД код: 25116003] – „Отговаря за удовлетворяване на изисквания-

та на клиента по отношение на потребителската визия, графичния интерфейс, интеракцията между приложението и потребителя, както и прилагането на стандартите и изискванията за ползваемост. Идентифицира евентуалните проблемни точки и дава предложения за подобряване на степента на ползваемост на приложението“.

И двата профила са в сектор „Информационни технологии“ Двете длъжности от Националната система за оценка на компетенциите разкриват сходни компетенции. И двете обаче не задоволяват нашето търсене, защото са ориентирани по-скоро към бизнес анализите.

Това наложи да се направи преглед и в системата на Европейската комисия за оценка на уменията, компетентностите, квалификациите и професиите (<https://ec.europa.eu/>). В резултат на това беше открит профилът на „изследователя на данни“ (Data Scientist), представен чрез необходимия набор от умения за реализация в професията. Таблица 3.3 разкрива задължителните и препоръчителните компетенции, включени в профила на специалиста по данни според ESCO.

Таблица 3.3. Компетенции на специалиста по данни според ESCO

Задължителни компетенции		Препоръчителни компетенции	
Основни умения и компетенции	Основни знания	Незадължителни умения и компетенции	Назадължителни знания
Боравене с извадки от данни; докладване на резултати от анализ; извършване на аналитични математически изчисления; почистване на данни; изграждане на системи за препоръчване; нормализиране на данни; предоставяне на визуална презентация на данни; прилагане на процеси за качество на данните;	Език за описание на търсене на мрежови ресурси; езици за запитвания; извличане на информация; категоризиране на информация; модели на данни; онлайн аналитична обработка; проучване на данни; статистика; техники за визуално представяне	Извършване на проучване на данни; интегриране на ИКТ данни; определяне на критерии за качество на данните; създаване на модели за данни; управление на архитектура с ИКТ данни; управление на данни; управление на класифициране на данни за ИКТ	LDAP LINQ MDX N1QL SPARQL Xquery Оценка на качеството на данните; неструктурирани данни; търговско разузнаване

Задължителни компетенции		Препоръчителни компетенции	
Основни умения и компетенции	Основни знания	Незадължителни умения и компетенции	Назадължителни знания
проектиране на схема за база данни; разработване на приложения за обработване на данни; събиране на ИКТ данни; тълкуване на текущи данни; управление на системи за събиране на данни; установяване на процеси за данни			

Специфичните компетенции в профила на специалиста по данни са представени в следващата таблица 3.4.

Таблица 3.4. Специфични компетенции на специалиста по данни

Компетенция	Описание	Задачи
Боравене с извадки от данни	Събиране и селектиране на набор от данни от популация чрез статистическа или друг вид процедура	Оценяване на данни, информация и цифрово съдържание
Докладване на резултати от проведен анализ	Изготвяне на документи или презентации с цел представяне на резултатите от проведен изследователски и аналитичен проект, като се посочват процедурите и методите за анализ, както и потенциалните интерпретации на резултатите	Докладване на факти, проверяване на данни и извършване на анализ на данни

Компетенция	Описание	Задачи
Извършване на аналитични математически изчисления	Прилагане на математически методи и изчислителни технологии, за извършване на анализи и решаване на проблеми	Извършване на маркшайдерски изчисления, на разчети за електрооборудване, на математически изчисления при борбата с вредители, на свързани с работата изчисления; прилагане на аналитично мислене
Извършване на почистване на данни	Откриване и коригиране на повредени записи; осигуряване на структурираност на данните съгласно изискванията	Прецизиране на данните
Нормализиране на данни	Редуциране на данните (нормални форми), за да се постигне минимизация на зависимостите; премахване на излишните данни; постигане на консистентност на данните	Установяване на процеси за данни
Предоставяне на визуална презентация на данни	Създаване на визуални изображения на данни, като графики и диаграми, с цел по-лесно разбиране	Прилагане на техники за визуално представяне на данни
Прилагане на процеси за качество на данните	Прилагане на техники за анализ на качеството, за валидиране и верификация	Използване на бази данни; прилагане на стандарти за качество; записване на производствените данни за контрол на качеството
Проектиране на схема за база данни	Проектиране на база данни с цел създаване на логически свързани групи обекти, като таблици, колони и процеси	Управление на база данни и създаване на инструменти за разработване на база данни
Разработване на приложения за обработване на данни	Създаване на персонализиран софтуер за обработка на данни чрез подходящ език за програмиране	

Компетенция	Описание	Задачи
Събиране на ИКТ данни	Прилагане на подходящи методи	Събиране на данни; управление на системи за събиране на данни
Интерпретиране на данни	Анализ на събраните от различни източници актуални данни – пазарни данни, научни доклади и изследвания, изисквания на клиентите и въпросници, с цел оценка на развитието и иновациите в разглежданата област	Тълкуване на сеизмични данни; разчитане на данни, свързани с производството на храни; тълкуване на данни за извличане; използване на методи за анализ на логистични данни; тълкуване на данни за автоматичното разпределяне на повикванията; оценяване на надеждността на данни; анализиране на данни за околната среда; тълкуване на геофизически данни
Управление на системи за събиране на данни	Разработване и управление на методи и стратегии, използвани за повишаване на качеството на данните и статистическата ефективност при събирането им, за да се гарантира, че събраните данни са оптимизирани за по-нататъшна обработка	Управление на данни; събиране на ИКТ данни; управление на колекции
Установяване на процеси за данни	Използване на ИКТ и инструменти за прилагане на математически алгоритмични или други процеси за обработка на данни с цел извличане на смислена информация от тях	Нормализиране на данни

3.2. ПРЕГЛЕД НА ПРОГРАМИТЕ В ОБЛАСТТА НА НАУКАТА ЗА ДАННИТЕ

3.2.1. Състояние и тенденции за развитие в световен мащаб

Възникващите технологии за бизнес интелигентни системи и социалните медии подхранват необходимостта от иновативни учебни програми, които да отговарят на нуждите на индустрията. Ясно е, че данните придобиват стойност след анализ, интерпретиране и внедряване в бизнес процесите. През последните години това създаде голямо търсене на специалисти по данни и бизнес анализатори. През последните 10 – 15 години организациите осъзнават ролята на специалистите по данни, затова назначават главен специалист по данни (Chief Data Officer), главен анализатор (Chief Analytics Officer), специалисти по моделиране на данни (data modelers) и анализатори на данни (data miners) в допълнение към длъжности, които се използват от по-рано, като инженери по данни (data engineers) и бизнес анализатори (business analysts).

Международната корпорация за данни (IDC) прогнозира необходимост от 181 000 нови позиции за специалисти със сериозни аналитични умения и от пет пъти повече работни места с нужда от умения за управление на данни и интерпретация до 2018 г. В стремежа си да отговорят на нарастващите потребности на различните организации от такива специалисти [60, 153, 189] много университети вече се ориентират в тази посока и започват да предлагат образователни степени в областта на науката за данните.

Науката за данните е сравнително нова академична област. В стремежа си да обхванат многостранните знания, които трябва да се преподават на бъдещите специалисти по данни, съгласно изследванията в научната литература, висшите училища разработват авангардни учебни програми или актуализират вече съществуващи такива в областта. Използват се педагогически иновации, различни казуси от практиката, упражнения с разнообразни софтуерни приложения, за да се включат най-съвременните инструменти и технологии в сферата на анализа и обработката на данни. През 2016 г. математическият институт Park City Mathematics Institute (PCMI) публикува ръководство за проектиране на бакалавърска програма по наука за данните, което е плод на съвместната работа на 25 факултета по компютърни науки, статистика и математика, обединили усилията си, за да изложат своето виждане за образованието в областта [227].

Вече има редица програми за придобиване на образователни степени в областта на науката за данните, стартирани от университети, училища или катедри и включващи бизнес, компютърни науки, статистика, математика, информационни системи и други науки [20, 209, 218]. В големите университети в Европа се предлагат и магистърски програми в областта. Например в Oxford University се предлагат две програми, свързани с науката за данните и извличането на знания от данни. В University of Cambridge има магистърска програма, а също и докторска, в областта на машинното обучение и технологиите (http://mlg.eng.cam.ac.uk/?page_id=659). Imperial College London предлага няколко магистърски програми – по машинно обучение, анализ на данни в биомедицината, статистика, а също и такава, ориентирана към бизнес анализи. Популярна инициатива през последните години в Европа е магистърската програма в областта на науката за данните, предлагана от EIT Digital (<https://masterschool.eitdigital.eu/programmes/dsc/>).

Друга инициатива на европейско равнище е European Data Science Academy (EDSA), която е част от европейски проект по програма H202 (2015 – 2018). Тя се превърна в онлайн институт, който продължава дейността си и след приключването на проекта. Съвместна инициатива на Scuola Normale Superiore, University of Pisa, Sant'Anna School of Advanced Studies и IMT School for Advanced предлага докторска програма в областта на анализа на данни (<https://datasciencephd.eu/>). Един от най-пълните и непрекъснато актуализирани списъци на академичните програми в тази област е предоставен от Data Science Community. Списъкът наброява повече от 600 образователни институции, предлагащи специалността „Наука за данните“. Появяват се все повече възможности за обучение чрез отворени онлайн курсове (Massive Open Online Courses, MOOC), насочени към широка аудитория обучаеми, които се предлагат както от академии, така и от университети. Подобна форма на обучение предлага University of Warwick, а курсът носи името „Big Data: Measuring and Predicting Human Behaviour“. Johns Hopkins University в САЩ предлага курс по наука за данните, а университетът Stanford предлага MOOC в областта на машинното обучение.

Zhang прави подробно проучване за образованието в областта на науката за данните в Китай от гледна точка на анализа на данни в библиотечните информационни системи [242].

Song and Zhu (2016) проучват 48 програми по науката за данните в САЩ, разделени в четири категории: бакалавърски, магистърски,

докторски програми, както и специализирани програми за получаване на сертификата на експертно ниво [209]. Повечето от тези програми са разработени чрез сътрудничество между множество обучаващи институции с оглед на мултидисциплинарния характер на областта и многообразието от знания и умения на специалистите по данни. Бакалавърски програми в областта на науката за данните все още се предлагат в малък брой университети, например в Ohio State University. Докторски програми почти не се срещат. Например University of Washington (<http://data.washington.edu/education/IGERT/index.html>) предлага докторска програма за обработка на големи данни, свързани с области като астрономия, инженерство и компютърни науки. Програми за получаване на сертификата обикновено се предлагат като онлайн курсове с продължителност, по-малка от една година, например в Columbia University (<http://datascience.columbia.edu/certification>).

Магистърските програми са най-често срещаните от четирите разглеждани категории. В Таблица 3.5 са посочени данни от проучването на Song and Zhu (2016), които се фокусират върху технически ориентирани програми в областта на науката за данните и анализа на данни и изключват области като бизнес анализ и управленски информационни системи [209].

Таблица 3.5. Основни курсове в програмите по наука за данните

Курс	Брой университети, които го предлагат
Бакалавърски програми	
Вероятности и статистика	7
Извличане на знания от данни (Data Mining)	7
Програмиране	5
Дискретна математика	4
Структури от данни и алгоритми	4
Бази от данни	4
Машинно обучение	4
Статистическо моделиране	3
Визуализация на данни	3
Въведение в науката за данните	2
Изкуствен интелект	2
Сигурност на компютърните системи	2
Магистърски програми	
Проучвателен анализ на данни (<i>Exploratory Data Analysis</i>)	10
Бази от данни	10

Курс	Брой университети, които го предлагат
Извличане на знания от данни	9
Визуализация на данни	8
Статистическо моделиране	8
Машинно обучение	6
Обработка и извличане на информация (<i>Information Retrieval</i>)	5
Анализ на информация и социални мрежи	4
Хранилища за данни	4
Въведение в науката за данните	3
Изследователски методи	3
Социални аспекти на науката за данните	3
Алгоритми	2
Почистване на данни (<i>Data Cleaning</i>)	2
Извличане на знания от текст	2
Анализ на данни в медицината	2

Както стана ясно в Глава II, науката за данните е обобщаваща рамка на много дисциплини, а академичните програми (бакалавърски, магистърски и докторски) в свързаните с нея области, като статистика, бизнес анализ, изкуствен интелект и машинно обучение се предлагат от по-отдавна [92]. Структурата и учебните планове на по-голямата част от академичните програми по наука за данните в света са сравнително близки – всички те потвърждават интердисциплинарния и синергетичен характер на науката за данните. Голяма част от дисциплините в учебните планове са фокусирани върху разширяването на математическите знания (линейна алгебра, теория на вероятностите, статистика и др.). Друга група дисциплини поставят акцент върху софтуерното инженерство (бази данни, управление на бази данни и свързаната с тях софтуерна инфраструктура). Основната група дисциплини (включително машинно обучение, *Deep Learning*), осигуряват компетенции за работа с конкретни технологии, необходими на специалистта по данни. В програмите за обучение по наука за данните участват и група бизнес ориентирани дисциплини, като маркетингов анализ, обработка на естествен език, биоинформатика и др. Те формират теоретични знания и практически умения, свързани с ефективното приложение и решаване на конкретни бизнес задачи и казуси в организациите. През последните години в тази група се включват и дисциплини, свързани с етичните проблеми в контекста на обработката, анализа и интерпретацията на данни.

В същото време наборите дисциплини, заложили в учебните планове на различните обучаващи организации и университети, се различа-

ват помежду си. Според Yap това разминаване се дължи на факта, че все още няма консенсус по отношение на обхвата на науката за данните [238]. В някои от разгледаните случаи учебната програма се съсредоточава върху жизнения цикъл на науката за данните, в други се базира на компетенциите, търсени от страна на бизнес сектора, а в трети се основава на решаване на проблеми – учебни казуси, свързани с различни приложни аспекти на науката за данните [115, 171, 172, 238].

Развитието на компетентности в областта на науката за данните понастоящем е трудна задача главно поради липсата на цялостна рамка, която може да бъде посочена като насока за уменията, изисквани от ИТ сектора.

Cleveland посочва шест области, в които трябва да се съсредоточат учебната програма за обучение на специалисти по данни, като прави и разпределение на процента на курсовете във всяка от областите, както следва [59]:

- мултидисциплинарни проучвания: анализ на данни от различни области – 25%;
- модели и методи за данни: статистически модели; методи за изграждане на модел; методи за оценка и разпределение въз основа на вероятностни изводи – 20%;
- изчислителна техника: хардуер; софтуер; изчислителни алгоритми – 15%;
- педагогика: планиране на учебната програма и подходи за преподаване – 15%;
- оценка на инструментите: проучвания на инструментите, използвани в практиката, на нуждите от нови инструменти и процесите за разработване на нови инструменти – 5%;
- теория: основи на науката за данните; основни подходи и математически изследвания за модели, методи и изчисления с данни, преподаване и оценка на инструментите – 20%.

Според Donoho в областта на науката за данните може да се разграничат следните по-важни дейности, които би трябвало да са в основата на обучението на бъдещите специалисти по данни [85]:

1. изследване и подготовка на данните;
2. представяне и преобразуване на данните;
3. изчисления с данни;
4. моделиране с данни;
5. визуализиране и представяне на данните;
6. наука за науката за данните.

Всяка от тези дейности е важна и зависима от другите. Често ня-

кои от дейностите се подценяват, например процесите в дейност (1) по подготовка и почистване на събраните големи данни. За други пък има вече създадени традиции в преподаването, като курсове по статистика и машинно обучение, които са част от дейност (4). От друга страна, в процесите на моделиране възниква необходимост от съчетаване на информация от изследваните данни с информация от външни източници в конкретна област в контекста на решаваните задачи, което усложнява тези дейности.

Според Hicks основните практически принципи при реализиране на обучение в областта на науката за данните и формиране на компетенциите на подготвяните специалисти по данни трябва да бъдат [122]:

- организиране на учебната програма около набор от разнообразни казуси;
- интегриране на компютърни изчисления и програмиране във всеки аспект на курса на обучение;
- развитие на абстрактно мислене, като същевременно се минимизира зависимостта от използване на математически означения;
- структуриране на дейностите в дисциплините като реалистична мимикрия на реален опит на анализ на данни;
- демонстриране на важността на критичното мислене и скептицизма чрез примери.

Song and Zhu предлагат няколко основни ядра, които трябва да бъдат в основата на образователните програми в областта на науката за данните, като според тях фокусът на обучението може да бъде поставен върху едно или повече от тях [209].

• **CDO (*Chief Data Officer*) дисциплини.** Фактът, че CDO е специалист по данни от високо ниво, който управлява проекти в областта на големите данни, предполага развиване на следните умения и компетенции: умения за комуникация и лидерство, бизнес проникателност, управление на проекти, процеси, техники и технологии за големи данни, управление на данни, системно мислене.

• **Дисциплини, свързани с жизнения цикъл на данните.** В основата им стои процесът на извличане на полезно знание от данните (*Cross Industry Standard Process for Data Mining, CRISP-DM*) [49] – ориентиране в спецификата на областта, разбиране на данните, подготовка на данните, моделиране, оценка и използване.

• **Дисциплини, свързани с технологиите за големи данни и техники за изграждане на модели,** включително Hadoop и неговата екосистема (Hbase, Hive, Pig, Mahout, Storm, Spark), NoSQL бази данни, облачни изчисления, визуализация и др.

- **Дисциплини, свързани с изследователските методи, използвани в анализа на данни, които изграждат критично и научно мислене.** Работата с големи данни по същество е свързана с извличане на полезно знание и прозрения, затова е необходимо обучаемите да се научат да формулират изследователски въпроси, да подхождат правилно към въпроса и да валидират резултатите.

- **Дисциплини, свързани с анализ на малки данни.** Работата с малки данни не включва задължително технологии като Hadoop и NoSQL, а работа с фундаментални техники за обработка на данни, като статистика, интелигентен анализ на данни, машинно обучение и визуализация.

- **Дисциплини от други отдели/катедри/факултети,** пряко свързани с мултидисциплинарния характер на работата с данни в зависимост от спецификата на съответната образователна програма.

- **Осигуряване на опит с реални проекти в областта.**

- **Осигуряване на сътрудничество на образователната институция с организации и правителствени структури,** които са източници на реални проблеми и предизвикателства с (големи) данни – те могат да подпомогнат обучаемите чрез осигуряване на софтуер, хардуер, стажантски програми, обучения и др.

- **Активно използване на отворени онлайн курсове.** Появяват се все повече възможности за обучение от академии и университети, насочени към широка аудитория обучаеми. University of Warwick предлага курс „Big Data: Measuring and Predicting Human Behaviour“. В Stanford се МООС в областта на машинното обучение. Много университети предлагат курсове по науката за данните и свързаните области в Coursera (<https://www.coursera.org/>).

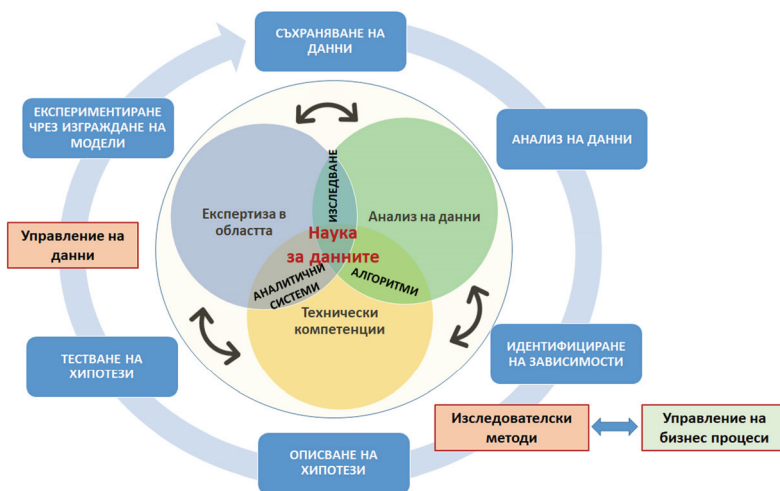
Проектът EDISON (Education for Data Intensive Science to Open New Science Frontiers) изгражда цялостна визия на сравнително новата на европейския пазар професия „Специалист по данни“. Основните компоненти на тяхната рамка (EDISON Data Science Framework, EDSF) за тази професия са:

- набор от компетенции (*Data Science Competence Framework*);
- основно ядро от знания (*Data Science Body of Knowledge*);
- модел на учебен план (*Data Science Model Curriculum*);
- професионален профил на специалиста по данни (*Data Science Professional profiles and Occupations Taxonomy*);
- таксономия на дисциплините в областта на науката за данните (*Data Science Taxonomy and Scientific Disciplines Classification*).

Наборът от компетенции на специалиста по данни определя основното ядро от знания, както и учебния план, който е в основата на обучението на бъдещите професионалисти. Рамката от основни компетенции на специалиста по данни е изработена в съответствие с Европейската рамка за електронна компетентност (European e-Competence Framework, e-CF3.0) и с предложение за нейното актуализиране и разширяване относно компетенции и умения в областта на науката за данните.

На Фиг. 3.2 са представени основните групи компетенции и връзката между тях:

- анализ на данни, включително статистически методи, машинно обучение и бизнес анализ;
- инженеринг: софтуер и инфраструктура;
- компетенции и знания в специфична научна област;
- управление на данни, организиране, съхраняване и сигурност;
- научни или изследователски методи (за научноизследователски профил) и управление на бизнес процеси (за бизнес ориентиран профил).



Фиг. 3.2. Групи компетентности в областта на науката за данните в образователни програми с общообразователен и изследователски профил (датирано от Demchenko [76])

На Фиг. 3.3 методите за управление на данни и за научни изследвания (или управление на бизнес процеси за втория профил) са поставени като два външни кръга, за да се подчертае, че тези компетенции и знания се изискват от всички професионални профили на специалистите. Препоръчва се във всички учебни програми по науката за данните да бъдат включени както *Управление на данни* (или *Research Data Management*), така и *Методи за изследване*.



Фиг. 3.3. Групи компетентности в областта на науката за данните в образователни програми с бизнес профил (датирано от Demechenko [77])

С оглед на посочените по-горе компетенции, които трябва да притежава специалистът по данни, в проекта EDISON се определят основните ядра от знания, необходими на професионалистите за извършване на всички процеси, свързани с обработката и анализа на данни. Тези ядра от знания представляват групи дисциплини, които имат за цел да изградят основните групи умения и компетенции и да служат като ориентир при изграждането на учебен план за обучение на такива специалисти. Предлаганият учебен план предоставя два основни компонента за изграждане на адаптивни учебни програми в областта на науката за данните:

- (1) дефиниране на резултатите от обучението въз основа на компетенциите, включително тяхното разграничаване за различни нива на владеене, с използване например на таксономията на Bloom;
- (2) дефиниране на учебни единици, които отразяват учебните

цели за целевата група. Те трябва да се определят в зависимост от съществуващите академични дисциплини, класифицирани според поръките на ACM/IEEE.

Рамката EDISON обхваща основни моменти, свързани с обучението на бъдещите специалисти по данни. Тя се базира на компетентностния подход и може да бъде използвана като концептуална основа за образование и обучение, професионално сертифициране, управление на организационни и индивидуални умения в областта на науката за данните.

3.2.2. Състояние на обучението по науката за данните в България

Връзката с компютърните науки, статистическите методи и модели, математическите изчисления и програмните езици ни дава основание да смятаме, че на национално равнище програмата за обучение по науката за данните трябва да бъде търсена в професионално направление 4.6. „Информатика и компютърни науки“. Принадлежността към друго професионално направление е допустима, но това безспорно би повлияло върху изграждането на задължителния комплекс от умения, включени в професионалния профил на специалиста по данни (виж Глава 2, раздел 2.2).

Прегледът в системата на Министерството на образованието и науката (<https://rsvu.mon.bg>) показва, че университетите в България, които обучават специалисти в областта на науката за данните, са 5: Софийският университет „Св. Климент Охридски“ (СУ), Техническият университет (ТУ), Нов български университет (НБУ), Варненският свободен университет „Черноризец Храбър“ (ВСУ) и Университетът за национално и световно стопанство (УНСС). Те предлагат магистърски програми за трисеместриално обучение в професионалните направления 4.6. „Информатика и компютърни науки“ и 3.8. „Икономика“, които са поместени в

Таблица 3.6. В България не бяха открити бакалавърски програми по наука за данните, единствено Софтуерният университет SoftUni предлага курс, завършващ със сертифициращ изпит. В Американския университет в България се предлага концентрация по Data Science в рамките на специалността „Информационни системи“.

Таблица 3.6. Национални магистърски програми в областта на науката за данните

Университет	ПН	Специалност	Магистърска програма
СУ	4.6	Информатика	Извличане на информация и откриване на знания
ТУ		Информатика и компютърни науки	Анализ на големи масиви и потоци от данни
НБУ		Информатика и компютърни науки	Извличане на знания и технологии за големи данни
ВСУ		Компютърни науки	Data Science (наука за данните)
УНСС	3.8	Бизнес информатика	Бизнес информатика, специалност „Информационни технологии за бизнес анализи“
СУ		Икономика	Моделиране на големи данни в бизнеса и финансите (Business Analytics) Счетоводство и анализ на големи данни – за неикономисти

Справка в Националната агенция за оценяване и акредитация (НАОА) показва, че на национално равнище е акредитиран само един университет, който да обучава в образователна и научна степен „Доктор“ в областта на науката за данните. Обучаващо звено по докторската програма „Аналитични изследвания върху данни“ (Data Science) е катедрата „Статистика и иконометрия“, а основно звено е Стопанският факултет на Софийския университет. Докторската програма е акредитирана в професионално направление 3.8. „Икономика“ с период на валидност до 2022 г.

С цел определяне на особеностите на програмите спрямо професионалния профил на специалиста по данни, в Таблица 3.7 са посочени основните изучавани дисциплини.

Специалностите в професионално направление 4.6 се доближават в по-голяма степен до компетентностния модел „твърди – меки – аналитични умения“. От друга страна, програмите от ПН 3.8 са ориентирани в по-голяма степен към изграждането на комуникационни и бизнес компетенции (т.нар. „меки умения“).

Широко застъпени в обучението по науката за данните в професионално направление 4.6. на национално равнище са дисциплини, изграждащи основно „твърди умения“ (Таблица 3.8).

Таблица 3.7. Дисциплини, изразждащи учебните планове на магистърските програми в областта на науката за данните, предложени на национално равнище

Дисциплина/Университет	ПН 4.6. „Информатика и компютърни науки“				ПН 3.8. „Икономика“		
	ВСУ	ТУ	НБУ	СУ	УНСС	СУ(1)	СУ(2)
МАТЕМАТИКА							
Вероятности и статистика	✓		✓		✓		
Матрични изчисления			✓				
Изчислителна геометрия				✓			
Размити множества и приложения				✓			
Високопроизводителни изчисления	✓		✓				
АРХИТЕКТУРИ И МРЕЖИ							
Компютърни архитектури и операционни системи	✓			✓	✓		
Компютърни мрежи и комуникации	✓		✓				
Компютърно моделиране на данни						✓	
Мрежа за съхраняване на данни							
Разпределени и облачни технологии и изчисления	✓	✓	✓	✓			
Уебтехнологии				✓			
ПРОГРАМИРАНЕ							
Програмиране	✓		✓		✓	✓	
Паралелно програмиране			✓				
Програмиране за IoT	✓						
Програмиране за Data Science	✓						
ТЕХНОЛОГИИ ЗА ДАННИ							
Бази данни	✓	✓	✓	✓	✓	✓	
Структури от данни и алгоритми	□	□	□	□			
АНАЛИЗ НА ДАННИ							
Качество на данните	□		□	□			□
Моделиране, оптимизация и визуализиране на данни		□	□		□	□	
Моделиране и управление на знания		□	□	□		□	
Методи за анализ на големи данни	□	□	□	□	□		
Алгоритми за извличане на данни	□		□	□	□		
Прогнозиране чрез анализ на данни			□			□	
Социалномрежови анализ	□						
ИЗКУСТВЕН ИНТЕЛЕКТ И ЕКСПЕРТНИ СИСТЕМИ							
Невронаука и невронни мрежи	□		□				
Машинно самообучение	□			□			
Изкуствен интелект; Интелигентни системи		□		□			

Дисциплина/Университет	ПН 4.6. „Информатика и компютърни науки“				ПН 3.8. „Икономика“		
	ВСУ	ТУ	НБУ	СУ	УНСС	СУ(1)	СУ(2)
Семантични данни и технологии	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>			
Обработка на естествен език и откриване на знания в текст	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>			
СИГУРНОСТ							
Криптография	<input type="checkbox"/>						
ИТ сигурност	<input type="checkbox"/>						
ПРОЕКТИРАНЕ НА ПРИЛОЖЕНИЯ И СИСТЕМИ							
Изграждане на приложения за големи данни			<input type="checkbox"/>				
Проектиране на софтуерни системи				<input type="checkbox"/>	<input type="checkbox"/>		
КОМУНИКАЦИОННИ И БИЗНЕС УМЕНИЯ							
Информационен мениджмънт	<input type="checkbox"/>						
Системи за управление на информацията	<input type="checkbox"/>						
Технологично предприемачество и иновации		<input type="checkbox"/>		<input type="checkbox"/>			
Професионална етика				<input type="checkbox"/>			
Управление на проекти				<input type="checkbox"/>		<input type="checkbox"/>	
Макроикономика					<input type="checkbox"/>		
Бизнес комуникации и бизнес анализ					<input type="checkbox"/>	<input type="checkbox"/>	
Информационна сигурност					<input type="checkbox"/>		
Бизнес разузнаване						<input type="checkbox"/>	
Финансов инженеринг						<input type="checkbox"/>	

Таблица 3.8. Дисциплини, изграждащи „твърди“ умения

Дисциплина	Честота на срещане
Вероятности и статистика	50%
Високопроизводителни изчисления	50%
Компютърни архитектури и операционни системи	50%
Компютърни мрежи и комуникации	50%
Разпределени и облачни технологии и изчисления	100%
Програмни езици	50%
Бази данни	100%
Структури от данни и алгоритми	100%
Качество на данните	75%
Моделиране, оптимизация и визуализиране на данни	50%
Моделиране и управление на знания	75%
Методи за анализ на големи данни	100%
Алгоритми за извличане на данни	75%
Невронаука и невронни мрежи	50%
Машинно обучение	50%
Изкуствен интелект; Интелигентни системи	50%
Семантични данни и технологии	75%
Обработка на естествен език и откриване на знания в текст	75%

Основното предизвикателство пред висшето образование в нашата страна, както и в глобален мащаб, е свързано с качеството на обучението и подготовката на компетентни и конкурентоспособни специалисти, отговарящи на изискванията на организациите. Университетите залагат на компетентностния подход, при който фокусът на образователния процес е насочен към изграждане на умения и компетенции, свързани с формирането на специалисти, които да подпомагат организациите в процесите на анализ на данни и вземане на информирани решения.

След проучване на компетентностите и изискванията на организациите, представено в Глава II, и на набора от изучавани дисциплини в учебните планове по науката за данните у нас и в чужбина става ясно, че съвременните образователни програми, подготвящи специалисти по една от най-търсените специалности в световен мащаб, днес успяват да осигурят едва половината от необходимите за реализацията на тези кадри знания и умения. Това поставя пред университетите задачата да разработят нови учебни програми и планове, като напускат зоната на тясно професионалното образование и приложат иновативни подходи както при разработването на учебното съдържание, така и в педагогическите инструменти, използвани при предлагането му.

3.3. ПРОУЧВАНЕ НА ВХОДЯЩИТЕ УМЕНИЯ НА ПОТЕНЦИАЛНИТЕ КАНДИДАТИ ЗА МАГИСТЪРСКА ПРОГРАМА ПО НАУКА ЗА ДАННИТЕ

3.3.1. Аналитичното мислене като основна компетентност

Основната задача на науката за данните е свързана с използване на съвременните технологии и инструменти за анализ на данни с цел извличане на полезно знание и решаване на различни проблеми [59]. Затова аналитичните компетенции на специалистите по данни са ключов елемент в техния професионален профил. В Глава II показваме, че аналитичните компетенции на специалистите, боравещи с данни, представляват пресечна точка на други две групи умения – „твърди“ (технически) и „меки“ (нетехнически, например комуникация, сътрудничество, любознателност и др.) умения, особено в контекста на големите данни. Освен това дефинирахме аналитичните умения като отделна група умения, за да подчертаем тяхното особено голямо значение за дейностите, които извършват специалистите във

вързка с жизнения цикъл на данните. Способността да се мисли логично, аналитично и критично, е много полезна в дейности като вземане на решения, анализ на въпроси, проучване, решаване на проблеми и самооценка – дейности, които са част от работата на специалистите по данни.

Развиването на аналитично и критично мислене е основна мисия на всички образователни институции. От студентите също се очаква да развият аналитично, логическо и критично мислене в хода на образователния процес. Много изследователи, преподаватели и обучаващи организации включват аналитичното и критичното мислене като основна компетентност при оценката на общите резултати от обучението на студенти [138, 146, 179, 183, 198, 217, 218]. Това означава, че съдържанието, структурата и последователността на учебните програми в учебните планове трябва да подпомагат развитието на аналитичното мислене у студентите.

На базата на гореизложеното ние предполагахме, че обучаемите, които притежават умения за аналитично мислене, ще се справят успешно в сферата на науката за данните. Следователно оценката на аналитичното мислене може да послужи като индикатор за студентите, които завършват различни бакалавърски програми, за да продължат обучението си в магистърска програма по наука за данните. От една страна, ще разглеждаме аналитичното мислене като индикатор за готовността на студента да завърши програмата успешно, а от друга – оценката на фона на потенциалните студенти играе решаваща роля в проектирането на отделните компоненти на магистърската програма: подбор на курсове, балансиране между теория и практика и избор на специфични педагогически техники за мотивиране на студентите в процеса на усвояване на знания и умения. Във връзка с това възникват следните въпроси:

1. Дали потенциалните студенти притежават необходимите аналитични умения за изучаване на магистърска програма по наука за данните?

2. Можем ли да предвидим потенциалния успех, който обучаващите могат да постигнат като специалисти по данни, когато завършат програмата?

3. Можем ли да подобрим магистърската програма по наука за данните, за да постигнем адекватни образователни модели и развитие на аналитичното мислене на обучаваните?

С цел проучване на настоящите студенти в бакалавърските програми в областта на информационните технологии, които ние разглеждаме като потенциални студенти в магистърска програма по

наука за данните, беше разработен въпросник за оценка на аналитичното им мислене. В следващите подточки са изложени нашият подход при съставянето на въпросника и резултатите от изследванията.

3.3.2. Разработване на въпросник

Процесът на разработване на въпросника за анализ на аналитичните умения (ВААУ) е представен на Фиг. 3.4.



Фиг. 3.4. Процесът на разработване на въпросника

Ще разгледаме поотделно всички етапи.

Дефиниране на целта на въпросника: Целта на разработването на този въпросник е оценка на аналитичните умения на респондентите. Въпросите са подбрани така, че с тях да се изследват разнообразни умения за разбиране, интерпретиране, формулиране на извод и решаване на различни проблеми и задачи.

Дефиниране на целевата група: Проучването е адресирано към аналитичните способности на настоящи студенти на Университета по библиотекознание и информационни технологии (УниБИТ), за които смятаме, че са потенциални кандидати за магистърска програма по наука за данните. Основната целева група бяха студенти от бакалавърски програми на специалностите „Компютърни науки“, „Информационни технологии“ и „Информационно брокерство“, които са преминали обучение относно различни аспекти на ИТ, но бяха включени и по-малък брой студенти от други, нетехнически, области, като „Библиотекознание“ и „Обществени комуникации“.

Изследване на съществуващи тестове за оценка на аналитичните умения:

тичното и критичното мислене: Ние изследвахме принципите и подходите, свързани с уменията за аналитично и критично мислене, както и съществуващите тестове за оценка на аналитичното и критичното мислене, тестове за умения за решаване на проблеми и др.

Таксономията на Bloom в по-високите нива на мисловните умения се използва за анализ и оценка на логическото и критичното мислене. В същото време повечето тестове са по-структурирани и обективни от таксономията на Bloom.

Популярна форма за оценка на критичното мислене е Калифорнийският тест за умения за критично мислене (*California Critical Thinking Skills Test, CCTST*). Той е насочен към основните умения за критично мислене: анализ, интерпретация, извод, оценка и обяснение [97]. Тестът генерира няколко резултата: общ резултат; резултат, свързан с оценката за индуктивни и дедуктивни разсъждения; резултат, свързан с категориите за анализ, извод и оценка [97]. CCTST обикновено се използва за оценка на резултатите от обучението, за оценка на учебните програми и плановете за професионално развитие и обучение и като елемент в процеса на кандидатстване и оценка на персонала.

За оценка на критичното мислене се използва тестът на Watson-Glaser (*Watson-Glaser Critical Thinking Appraisal, WGCTA*), концентриран върху уменията за формулиране на извод, индукция, дедукция, предположения, интерпретация и оценка на аргументите. Тестът на WGCTA, използва само един резултат [234] и се прилага основно за ученици и студенти.

Други подобни съществуващи тестове за оценка на критичното мислене са *Cornell Critical Thinking Tests*, *the Ennis-Weir Critical Thinking Essay Test*, *New Jersey Test of Reasoning Skills* и *Ross Test of Higher Cognitive Processes* [90, 91]. Тестът на Cornell (*Cornell Critical Thinking Tests*) има две нива, X и Z, които са насочени към различни възрастови групи. Състои се основно от затворени въпроси с множествен избор, които изследват важни аспекти, като способностите за индуктивен извод, разкриване на логически заблуди и достоверност, прогнозиране и експериментално планиране, формулиране на предположения и др. Тестът на Ross (*Ross Test of Higher Cognitive Processes*) оценява високите нива на мисловните умения според таксономията на Bloom, а именно способностите на студентите да извършват анализ, синтез и оценка. Състои се от 105 затворени въпроса с множествен избор. Тестът на Ennis-Weir (*Ennis-Weir Critical Thinking Essay Test*) е насочен към обучаеми в колежи и университети, както и към други възрастни. Много прилича на теста на Cornell по аспектите, които засяга, но основната разлика между двата теста е, че тес-

тът на Ennis-Weir се предлага под формата на есе.

Съществуващите тестове, които се изследват в научната литература в тази област, варират в широки граници както по предназначение, така и по формат, но обикновено целят получаване на глобална оценка на критичното мислене като характеристика на индивида, а не в определен контекст или специфична област. Тестовите съдържат задачи, които обхващат важни аспекти на количественото, логическото, критичното и аналитичното мислене.

Съществуват редица предизвикателства при оценката на уменията за критично мислене у обучаваните. Според Ennis тестовите за оценяване на критичното мислене не са изчерпателни, особено тези с множествен избор, които са и най-лесни за използване. Съществуващите тестове не успяват да обхванат важни компоненти на мисловния процес, като възприемчивост, способност за преценяване на достоверността на източниците, размисъл над алтернативи, формулиране на хипотези и дефиниции и разработване на план за провеждане на експерименти. Ennis смята, че съществува голяма разлика в начина на оценяване на тестовите с въпроси с множествен избор и на тестовите с кратък отговор или есе.

Разработване на концептуална рамка на въпросника: Тестът цели да разкрие количествените и качествените аспекти на формирането на аналитично мислене в процеса на обучение на студентите. Ето защо, като се вземат предвид проучената научна литература, включително проучените аспекти на критичното и аналитичното мислене, съществуващите тестове за оценка на аналитичното/критичното мислене, целта на въпросника и целевите респонденти, се появява необходимостта от рамка, която да определя измеренията на изследването. На този етап е целесъобразно да се извърши методичен преглед на:

- броя на въпросите (за да се гарантира, че респондентите няма да се отегчат бързо – 20 бр.);
- методите за оценка на получените резултати;
- крайните срокове (за изготвяне на въпросника, събиране и анализиране на данните);
- методите за достигане до респондентите (хартиен вариант на въпросника или онлайн тест);
- възможните формати на изследването (лично интервю, анкета).

Определяне на категориите от въпроси: Въвеждащите въпроси в анкетата имат за цел да подтикнат студентите да направят оценка на своите умения за аналитично мислене.

Основната част от въпросника предлага въпроси с множествен избор в четири категории, както следва:

1. Въпроси за изследване на аналитичните умения (**АУ**), включително проверка на способността за идентифициране на прилики, разлики, правилна последователност, противоречия;

2. Въпроси за оценка на способностите за абстрактно мислене (**АМ**) (доколко в отговорите си респондентите следват дефинициите на обектите, предложени от въпросника, а не прилагат интуитивното, добре познато разбиране на техните свойства);

3. Въпроси за разграничаване на факти от мнения (**ФМ**);

4. Въпроси и задачи за изследване на способността за количествено мислене (**КМ**), включително сравняване на количества и прилагане на основни добре познати математически знания във всекидневни ситуации.

Определяне на формата на въпросите: Конструираниите въпроси представляват логически задачи в три различни формата: *математически въпроси*, *въпроси за работа с текст* и *въпроси с фигури*. Те са конфигурирани така, че респондентите да могат да избират от няколко възможности. Въпросите за самооценка на аналитичните способности на респондентите са от типа Likert скала. Използвахме тристепенна скала – *високо*, *средно* и *ниско ниво* на прилагане на аналитичните умения.

Въпросите бяха разпределени на случаен принцип между различните категории и формати.

Формулиране на въпросите: Някои елементи на въпросите са приети или адаптирани от съществуващи тестове, а други са нови, направени специално за този въпросник съгласно поставените изследователски задачи (вж. Приложение 1).

Адаптиране на въпросите към целевите респонденти: По време на разработването на въпросника анализирахме трудностите, които могат да окажат влияние на активното участие, мотивацията и прецизната работа на студентите от гледна точка на разбирането и интерпретирането на въпросите, техния брой, времетраенето на теста и т.н. Затова адаптирахме въпросите към вероятното ниво на аналитични умения на респондентите.

Избор на метод за събиране на данни: Проучването е планирано основно като онлайн въпросник.

Провеждане на пилотно проучване: С цел откриване на грешки и недостатъци във въпросника по отношение на съдържание (свръх детайлизиране, двусмисленост и др.), граматика и формат тестът беше предложен на наши колеги (университетски преподаватели)

и докторанти от катедрите „Компютърни науки“ и „Информационни системи и технологии“.

Провеждане на същинско проучване: Проучването е направено през есента на 2018 г. В него се включиха 250 респонденти.

3.3.3. Анализ на събраните от проучването данни

Целта на изследването е да бъде изготвена оценка на аналитичните умения, притежавани от респондентите. Проучването е адресирано към студенти, обучаващи се в (или завършващи) ОКС „Бакалавър“ в областта на информатиката и компютърните науки.

Събирането на данните чрез изготвения въпросник се осъществява както по електронен път, така и чрез проучване на терен. Изследването беше проведено на два етапа в рамките на сравнително близък времеви период.

При предварителната обработка бяха отстранени непоследователните или невалидни отговори. Бяха отчетени общо 235 попълнени въпросника. След валидирането им 32 от тях бяха отчетени като невалидни. Останалите 203 бяха използвани за анализ на проучването. Наличната извадка ни осигури валидна степен на отговор от 81,2% (Таблица 3.9).

Таблица 3.9. Разпределение на извадките

Извадка	Брой отчетени въпросници	Невалидни	Валидни
Първа	187	21	166
Втора	48	11	37
Общо	235	32	203

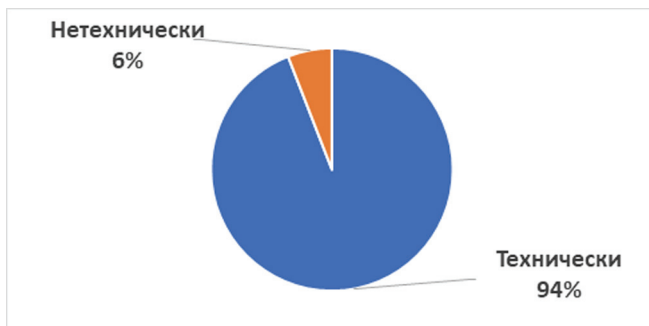
В изследването се включиха студенти от професионално направление „Информатика и компютърни науки“ (включително студенти от специалностите „Компютърни науки“, „Информационни технологии“, „Информационно брокерство“), както и студенти от професионални направления „Обществени комуникации и информационни науки“ и „Национална сигурност“. Процентното разпределение на студентите според принадлежността им към дадено професионално направление е поместено в Таблица 3.10. Данните показват, че малко над 94% от респондентите се обучават в областта на информатиката и компютърните науки, докато останалите около 6% са от нетехнически специалности.

Таблица 3.10. Разпределение на респондентите по професионално направление

Професионално направление	Брой	%
Информатика и компютърни науки	191	94,1
Обществени комуникации и информационни науки	7	3,4
Национална сигурност	3	1,5
Друго	2	1
Общо	203	100

Респондентите бяха разделени в две групи (фиг. 3.5):

- целева група (технически специалности) – включва студентите от специалностите „Информатика и компютърни науки“, „Информационни технологии“, „Компютърни науки“ и „Информационно брокерство“;
- контролна група (нетехнически специалности) – включва студентите от специалностите „Национална сигурност и културно-историческо наследство“, „Библиотекознание и библиография“ и „Комуникации и информирание“.



Фиг. 3.5. Процентно разпределение на респондентите от целевата и контролната група

Според събраните данни от въпросите за самооценка 22,1% от респондентите първоначално са оценили себе си като притежаващи най-висока степен на аналитично мислене. За сравнение 7,95% оценяват уменията си в края на теста с по-нисък резултат от първоначалния. Останалите респонденти запазват едни и същи стойности за оценка на своите умения както в началото, така и в края на теста (Таблица 3.11).

Таблица 3.11. Самооценка на аналитичните умения на техническата група

Направете самооценка на притежаваните от Вас аналитични и логически умения. *Как бихте оценили своите аналитични и логически умения след решаването на задачите от теста? Crosstabulation					
		22. Как бихте оценили своите аналитични и логически умения след решаването на задачите от теста?			Общо
		Ниска степен	Средна степен	Висока степен	
3. Направете самооценка на притежаваните от Вас аналитични и логически умения.	Висока степен	0	8	35	43
	Средна степен	14	61	39	114
	Ниска степен	8	9	2	19
Общо		22	78	76	176

Приложеният алгоритъм за обработка на аналитичните умения на респондентите включва 5 основни етапа: (1) рекодиране; (2) формиране на бал; (3) описателна статистика; (4) нормиране на скалите и (5) корелационен анализ.

Седемнайсетте въпроса, участващи във въпросника, оформиха четири подскали:

- АУ (въпроси 6, 7, 16, 17 и 21);
- АЛМ (въпроси 10, 11, 12 и 13);
- ФМ (въпроси 14 и 15);
- КМ (въпроси 5, 8, 9, 18, 19 и 20).

На първия етап на обработка на данните всичките 17 въпроса, участващи във формирането на четирите подскали, бяха рекодирани, като код 0 беше приписан на грешните, а код 4 – на верните отговори, съобразно ключа към картата.

В изпълнение на втория етап от алгоритъма за обработка на данните беше формиран бал по всяка от скалите като сума от отговорите на въпросите, които влизат в нея. Всеки респондент е описан чрез броя на правилните отговори във всяка категория въпроси. Например, ако студентът X отговори правилно на два въпроса в категорията АУ, на два в АМ, на един във ФМ и на три в КМ, студентът $X = \{AU(2), AM(2), FM(1), KM(3)\}$ може да се опише с авектора $X = \{2, 2, 1, 3\}$.

Като част от изпълнението на третия етап от алгоритъма бяха приложени техники за трансформация на данните, в резултат на което беше получена следната описателна статистика на отделните скали (Таблица 3.12).

Таблица 3.12. Описателна статистика на отделните скали

		Аналитични умения	Абстрактно-логическо мислене	Факти от мнения	Количествено мислене
N	Валидни	196	200	202	199
	Невалидни	8	4	2	5
Средна стойност		10,90	10,30	4,50	15,46
Стандартно отклонение		4,70	3,77	3,01	4,97
Минимална стойност		0	0	0	0
Максимална стойност		20	16	8	24

С цел да се диференцират лицата с ниско, средно и високо ниво на аналитични умения, беше извършено поотделно нормиране на оценката на бала за четирите подскали. Средното равнище на всяка скала се получава, когато балът е в интервала на средната стойност \pm стандартното отклонение. Съответно ниското равнище е под този интервал, а високото – над него. В резултат на приложеното нормиране бяха получени процентни разпределения, които ще бъдат представени в следващите четири таблици.

Таблица 3.13. Прилагане на аналитични умения

Аналитични умения					
		Честота	Процент	Валиден %	Кумулативен %
Валидни	Ниска степен	29	14,2	14,8	14,8
	Средна степен	111	54,4	56,6	71,4
	Висока степен	56	27,5	28,6	100,0
	Общо	196	96,1	100,0	
Невалидни	System	8	3,9		
Общо		204	100,0		

Таблица 3.14. Прилагане на абстрактно и логическо мислене

Абстрактно-логическо мислене					
		Честота	Процент	Валиден %	Кумулативен %
Валидни	Ниска степен	26	12,7	13,0	13,0
	Средна степен	142	69,6	71,0	84,0
	Висока степен	32	15,7	16,0	100,0
	Общо	200	98,0	100,0	
Невалидни	System	4	2,0		
Общо		204	100,0		

Таблица 3.15. Различаване на факти от мнения

Факти от мнения					
		Честота	Процент	Валиден %	Кумулативен %
Валидни	Ниска степен	46	22,5	22,8	22,8
	Средна степен	85	41,7	42,1	64,9
	Висока степен	71	34,8	35,1	100,0
	Общо	202	99,0	100,0	
Невалидни	System	2	1,0		
Общо		204	100,0		

Таблица 3.16. Прилагане на количествено мислене

Количествено мислене_					
		Честота	Процент	Валиден %	Кумулативен %
Валидни	Ниска степен	24	11,8	12,1	12,1
	Средна степен	161	78,9	80,9	93,0
	Висока степен	14	6,9	7,0	100,0
	Общо	199	97,5	100,0	
Невалидни	System	5	2,5		
Общо		204	100,0		

Обобщен анализ на данните, представящи степента на прилагане на уменията на респондентите за всяка от четирите скали, е предложен в следващата таблична форма.

Таблица 3.17. Прилагане на АУ, АЛМ, ФМ и КМ от целевата група

	Аналитични умения		Абстрактно-логическо мислене		Факти от мнения		Количествено мислене	
	Брой	%	Брой	%	Брой	%	Брой	%
Ниска степен	29	14,2	26	12,7	46	22,5	24	11,8
Средна степен	111	54,4	142	69,6	85	41,7	161	78,9
Висока степен	56	27,5	32	15,7	71	34,8	14	6,9
Всичко	196	96,1	200	98	202	99	199	97,5

Корелационният анализ, осъществен при изпълнението на последния етап от алгоритъма, показва, че е налице положителна и умерена по сила статистически значима взаимовръзка между категориите въпроси АУ АЛМ и КМ, от една страна, както и между АЛМ и КМ, от друга страна. ФМ не корелира статистически значимо с нито една подскала от изследването. Допускането е, че ако броят на въпросите

от категорията ФМ са повече два, някои тенденции във взаимовръзките между променливите ще имат статистически достоверен ефект (например между ФМ и АУ, където нивото на значимост на корелационния коефициент има гранични стойности – $r = 0,133$ при $p = 0,065$). Резултатите от корелационния анализ може да бъдат проследени в Таблица 3.18.

Таблица 3.18. Корелационен анализ

		Аналитични умения	Абстрактно-логическо мислене	Факти от мнения	Количествено мислене
Аналитични умения	Pearson	1	0,332*	0,133	0,368*
	Sig. (2-tailed)		0,000	0,065	0,000
	N		194	195	192
Абстрактно и логическо мислене	Pearson		1	0,061	0,296*
	Sig. (2-tailed)			0,388	0,000
	N			200	196
Факти от мнения	Pearson			1	0,100
	Sig. (2-tailed)				0,161
	N				198
Количествено мислене	Pearson n				1
	Sig. (2-tailed)				
	N				

* Correlation is significant at the 0,01 level (2-tailed).

РЕЗУЛТАТИ И ИЗВОДИ ПО ГЛАВА III

Основните изводи, които може да се направят след проучването на различните образователни стратегии и подходи при подготовката на специалисти в областта на науката за данните, са следните:

1. Техническите компетентности са важни, но не са критични за изграждането на успешна кариера на специалистите по данни. При тежаването на твърди умения е необходимо, но не достъчно условие за успех в тази област.

2. Аналитични умения и нагласа, ориентирана към решаване на проблеми, се възпитават трудно, но са от съществено значение. Меките умения са изключително важни за успеха при работа в екип и особено при работа в трансдисциплинарни екипи.

3. В различните образователни профили у нас акцентът се поставя преди всичко върху твърдите умения. Тези умения по-лесно се оценяват при търсене на служители и присъстват във всички обяви за работа.

4. Университетите имат богат опит в изграждането на твърди умения и естествено ориентират своите учебни програми в тази посока, подценявайки развитието на другите аспекти на подготовката на специалистите по данни.

Резултатите от проведената анкета показват, че и студенти с нетехнически профил биха могли да се изградят като добри специалисти по данни.

ГЛАВА IV.

ИЗГРАЖДАНЕ НА МОДЕЛ НА УЧЕБЕН ПЛАН, БАЗИРАН НА КОМПЕТЕНЦИИ

Създаването на специалисти е труден и продължителен процес, в който ролята на университетите е свързана с предлагането на модерно образование и формирането на кадри за конкретни сектори и отрасли на икономиката. Днес, в условията на динамични технологични промени, бизнесът очаква образователната система да реагира адекватно и да формира необходимите знания, компетенции и умения у адаптивни, иновативни и мотивирани специалисти, владеещи изкуството за работа в екип и учене през целия живот. Естеството на образователната система е консервативно и тя не може бързо да се пренастройва поради спецификата и продължителността на процесите от възникване на нуждата от специфично знание до неговото формиране у обучаемите. Вследствие на това нуждите на реалната икономика почти винаги изпреварват предложенията на пазара на труда и в частност кадрите, формирани в сферата на образованието.

Успехът на една програма за обучение зависи от това доколко тя позволява на обучаваните да навлизат в нея постепенно и лесно. Плавното въвеждане в контекста на конкретната изучавана област и насърчаването на студентите да се възползват от предлаганите предизвикателства, естествени за всяко ново начинание, са от съществено значение при разработването на нови учебни програми. Това е особено важно при проектирането на учебни програми за обучение в нова, неутвърдена област на образование.

В тази глава ще бъде предложен модел на учебна програма по наука за данните – една сравнително нова научна област, която включва вече утвърдени дисциплини, като статистика, програмиране, изкуствен интелект и др. Изграждането на учебната програма се базира на компетентностния модел и формирането на необходимите компетенции за бъдещите специалисти по данни.

4.1. ФОРМИРАНЕ НА КОМПЕТЕНЦИИ В ОБЛАСТТА НА НАУКАТА ЗА ДАННИТЕ

4.1.1. Методика за разработване на учебна програма, основана на компетенции

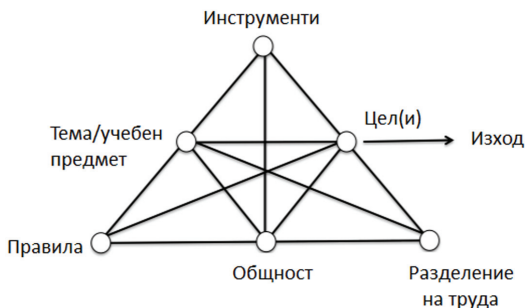
Образователните стандарти, установени в Европейското пространство за висше образование, предполагат, че целта на университетите е да дадат възможност на студентите да придобият компетенции в конкретна изучавана от тях професионална област. Фокусът на обучението, базирано на компетенции, е свързан както с придобиване на необходимите знания, съответстващи на професията, така и със способността придобитите знания да бъдат приложени в конкретен контекст. Най-общо казано, целта на проектирането на учебни програми, базирани на компетенции, е да се гарантира, че след като са придобили необходима комбинация от знания, умения и способности, обучаемите ще могат да демонстрират усвоените умения. Изграждането на подобен учебен план изисква сериозна теоретична подготовка, анализ и планиране.

Ясно формулираната теоретична основа при разработването на програма за обучение по наука за данните е от голямо значение, тъй като дава пълна картина на бъдещото развитие на обучението. От многото съществуващи различни теории на образованието, които би могло да се приложат при проектирането на обучителни курсове по наука за данните, най-подходяща е теорията на дейността (Activity Theory) [74, 142, 191]. Тази теория представлява хибрид от психологически факти и аксиоми на марксизма-ленинизма и господства в съветската психология до края на 80-те години. Основната теза в нея е, че „съзнанието определя дейността, а дейността определя съзнанието“.

Основната причина, поради която теорията на дейността е подходяща при проектирането на учебни програми по наука за данните, се свързва с контекста на големите данни:

- сурови данни;
- въпроси, произтичащи от суровите данни;
- съгласувани цели;
- трансформация на данните;
- използване на инструменти за анализ на данните.

Вследствие на извършено проучване за изграждането на учебен план, базиран на теорията на дейността, показва връзката с т.нар. „триъгълник на дейностите“ (Activity Triangle) [87, 88, 89, 186]. Примерна структура на този триъгълник е показана на следващата фигура.



Фиг. 4.1. Обща структура на триъгълника на дейностите (адаптирано от Yan, Donghui & Davis, Gary, 2019 [238])

Триъгълникът описва взаимовръзката между основните елементи в обучението по науката за данните:

- **Общност** – включва мнозинството от студенти и техните преподаватели (лектори). В частни случаи преподавателят в тази общност може да бъде заменен от лица, за които студентите трябва да създадат продукт от данни (като например прогнозен модел или доклад);

- **Тема/учебен предмет** – обхваща темата или предмета, изучаван/и от отделен студент или група от студенти, които работят заедно за постигането на един резултат (обикновено писмен доклад за анализа на данни);

- **Инструменти** (във връзка с обучението в областта на науката за данните):

- (1) *софтуерни инструменти* – тук се включват езици за програмиране, сред които R и Python;

- (2) *концептуални инструменти*, например регресия или техники за групиране. Студентите използват инструментите, за да постигнат поставената цел при решаване на конкретен казус;

- **Цел** – определя се чрез обсъждане между членовете на екипа (студентите), лектора и всеки външен клиент;

- **Правила** – варират в зависимост от дейностите и могат да включват някои общи такива, като: избягване на плагиатство; подходящо цитиране на източниците; сътрудничество в рамките на и между студентски екипи; споделяне на констатации; етично поведение и отговорност за спазване на крайните срокове;

- **Разделение на труда** – наблюдава се както на вътрешно равнище, между *самите студенти в екипа*, така и между *отделните екипи*, фокусиращи се върху различни аспекти на дадена цел.

Един и същ набор от данни обикновено позволява генерирането на редица различни дейности и цели, много често още по време на формулирането на допълнителни въпроси относно данните и на проучването на възможностите за постигане на определените цели. Когато това се случи, се променят всички елементи по веригата: ако се промени целта, ще се наложи да се променят и темите, така че студентите да формират нови групи (спонтанно или по указание на лектора). Това от своя страна ще промени и разделението на труда, а инструментите най-вероятно ще трябва да бъдат модифицирани и/или да се внесат нови, по-подходящи за постигането на целта.

Триъгълникът на дейността се координира с жизнения цикъл на науката за данните. Въпреки че са възможни промени, както е описано по-горе, някои аспекти на триъгълника на дейността остават фиксирани по време на цялото обучение, например:

- общността включва както потока от студенти, така и лекторите;
- правилата се формулират в началото на семестъра и остават повече или по-малко фиксирани и валидни до неговия край;
- обикновено разпределението на труда, в рамките на или между отделните групи, остава почти непроменено по време на цялото обучение.

Теорията на Yan [238], че напредъкът по даден проект включва преминаване през жизнения цикъл на науката за данните, е обяснена с това, че в рамките на своето обучение всеки студент преминава през жизнения цикъл на науката за данните, появява се нов триъгълник на дейността, в който новите въпроси формират нови цели, новите анализи изискват нови инструменти, а новите резултати водят до нови заключения. Така студентът постоянно се сблъсква с динамична последователност от триъгълници на различни дейности.

Теорията на дейността, приложена в обучението по науката за данните, изисква формиране на умения за трансформиране на данните чрез използване на наличните инструменти. През призмата на всичко казано дотук може да се обобщи, че данните се трансформират чрез активност, която се свързва с конкретна цел, а целта обикновено е резултат от естествено възникнал въпрос, свързан с наличните данни.

Възможностите на специалиста по данни да използва наличните инструменти за трансформиране и последващ анализ на данните, зависят не само от неговите знания, но и от уменията, които той прилага при решаването на конкретен проблем. Обучението на бъдещия специалист по данни трябва да бъде ориентирано към формиране на професионални компетенции, съобразени с търсените от страна на бизнес организациите.

Обучението, базирано на компетенции, от една страна, е важна стъпка в развитието на науката за данните, а от друга, ще помогне да бъде преодолявана съществуващата форма на дигитално разделение по повод на възможностите за използване на големи данни. Независимо коя теоретична рамка е приложена при изграждането на учебната програма, е важно като краен продукт тя да предполага обучение, базирано на компетенции.

Възприетият още в края на миналия век подход „компетенции за действие“ се характеризира с преминаването от изучаването на предмети към усвояването на компетенции и с разработването на нови учебни планове, определящи области на обучение, насочени към свързаните с бъдещата работа знания и умения.

Най-важната характерна черта на модела на обучение, базирано на компетенции, е фокусирането му върху учебните планове, в които трябва да се посочат:

- съвкупността от компетенции, специфични за дадения предмет и определящи като цяло приоритетните области на обучение;
- планираните за усвояване знания, умения и навици.

Стандартната типология на компетенциите е ориентирана към бъдещата професионална дейност на випускниците на учебните заведения [4]. Разработването на учебен план, основан на компетенции, започва с изграждането на успешна стратегия [169] и изисква сериозен литературен преглед на теоретичните основи на подхода.

Nikolov et al. разделят процеса на разработка на учебен план, базиран на компетенции, на две основни части [169]:

- първа част – съдържа пет стъпки, включващи оценка на готовността на университета за преминаване от традиционно съдържание на учебните програми към учебни програми, базирани на компетенции;
- втора част – описва етапите на превръщане на традиционния учебен план в план, базиран на компетенции.

В настоящия раздел ще предложим обобщена **систематична методология за разработване на учебна програма** (в частност учебен план на магистърска програма), базирана на компетенции. Методологията включва пет основни фази:

- *Концептуализация и планиране* – фаза, в която трябва да се уточни необходимостта от обучението (наложена едновременно от пазарното търсене и предлагането на специалисти по данни). В тази фаза е необходимо да се изяснят факторите, които определят нуждата от преход към обучение, формиращо професионални компетенции. За целта трябва да се осъществят събиране и анализ на данни, подпомагащи формирането на отговори за всеки от поставените в тази фаза въпроси.

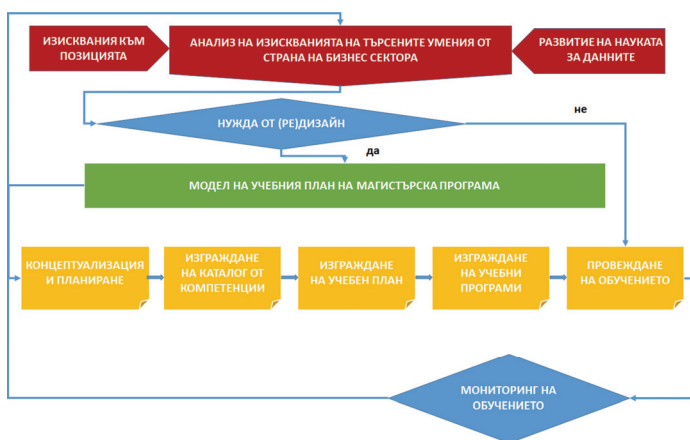
- *Създаване на каталог от компетенции* – идентифициране и описание на компетенциите, които студентите трябва да придобият по време на своето обучение. Тук трябва да бъдат описани най-високите нива на компетенции, заложиени в програмата.

- *Разработване на учебен план, основан на компетенции* – компетенциите, които трябва да бъдат включени в учебния план, се дефинират чрез изследване на съществуващите модели на компетенции в бизнес сектора и чрез изучаване на обявените работни позиции. Обикновено тази дейност се осъществява от хетерогенен екип, състоящ се от заинтересовани страни, като разработчици на учебни програми, учители, ръководители на обучението, експерти от сектора и представители на бранша. Всички те заедно анализират и събират информация за компетенциите в областта, идентифицират задължителните компетенции и ги описват в каталога от компетенции. Както става ясно, процесът на дефиниране на профила на компетенциите е много сложен и труден, има важни последици и служи като основа при създаването на учебна програма, базирана на компетенции.

- *Разработване на учебни програми* – разработването на учебни програми за всяка от включените в учебния план дисциплини е логично продължение на процеса на изграждане на магистърската програма.

- *Осъществяване на пилотен тест на обучението.*

Методиката на разработване на учебна програма, базирана на компетенции, е графично представена на Фиг. 4.2.



Фиг. 4.2. Систематичен подход за разработване и редизайн на магистърска програма, базирана на компетенции

Методиката насърчава непрекъснатото усъвършенстване чрез разглеждане на стратегическите планове на университета. Тя се основава на интензивното изследване и анализ на потребностите на разпитието и обучението, основано на компетенции.

Във връзка с направеното проучване в релевантната литература относно изграждането на обучение, базирано на компетенции, и по повод на проектирането на учебна програма по „Анализ и управление на данни“ се налагат следните обобщения:

- Разработваният модел на система за формиране на професионални компетенции за анализ на големи данни трябва да бъде андрагогически.

- Моделът трябва да бъде основан както на обхвата и достигнатото ниво на науката за данните в теоретичен аспект, така и на търсения набор от компетенции от бизнес организациите.

- Ефективното обучение на специалисти по данни трябва да бъде компетентностно базирано, личностно ориентирано, мотивиращо и непрекъснато усъвършенстващо се.

- За успешната реализация на завършващите студенти е необходимо обучението в областта на науката за данните да бъде ориентирано към изграждането както на знания и умения, така и на компетенции.

За изграждането на компетентностно ориентирано обучение по „Анализ и управление на данни“ е необходимо да бъде изграден общ модел на система за формиране на професионални компетенции.

4.1.2. Общ модел на система за формиране на професионални компетенции в областта на науката за данните

Моделът на професионални компетенции за анализ на големи данни трябва да служи като основна отправна рамка при всички изследователски процедури, включително:

- поставянето на целта и оценяването на компетенциите;
- проектирането на учебното съдържание и моделирането на взаимодействията в процеса на обучение;
- процедурите на контрол.

Моделът на професионалните компетенции включва четири основни компонента, всеки от които може да се разглежда като част от взаимно свързана подсистема [5]:

- *целово-резултативен компонент* – описва целите и очакваните резултати от обучението;

- *информационно-дидактически (съдържателен) компонент* – включва варианти на учебно съдържание и учебни задачи за неговото организирано усвояване;
- *процесуален компонент* – модел на процеса на обучение в аудиторна и извънаудиторна среда, включващ методите и техниките на обучение и организацията на учебния процес;
- *контрол и обратна връзка* – средствата, методите и техниките на контрол на резултатите, обратната връзка със субектите на процеса на обучение и вземаните решения за управление на поведението на системата.

Отделните компоненти и техните елементи са йерархично свързани. В първото ниво се посочват целите и очакваните резултати от обучението. Второто ниво представя учебното съдържание, третото – учебния процес, а четвъртото ниво отговаря за контрола и обратната връзка в курса на обучение. Нивата са свързани помежду си с канали за управление, осъществени чрез процедури за контрол и самоконтрол. Общият модел на системата е представен на Фиг. 4.3.



Фиг. 4.3. Общ модел на система за формиране на професионални компетенции [5]

Целево-результативният компонент на системата за формиране на професионални компетенции се свързва както с развитието на науката за данните, така и с потребностите на бизнес организациите в област-

та на анализа на големи данни. Освен това целево-резултативният компонент на системата може да се свърже и с характера на академичната подготовка на студентите в магистърска степен от специалност „Анализ и управление на данни“. Компонентът е образ на резултата, към който е ориентирана академичната подготовка на студентите, и включва [5]:

- усвояване на определен обем знания в областта на науката за данните;
- формиране на умения за анализ на големи данни на определено равнище;
- изграждане и закрепване на отношение и поведенчески нагласи, които формират професионалните компетенции за анализ на големи данни.

Елементите на подсистемата на образователните цели за академичната подготовка на студентите от специалност „Анализ и управление на данни“ са дефинирани според таксономията на Bloom [41], като се спазват общите изисквания всяка компетенция и нейните компоненти да бъдат измерими и проследими.

За модела на система за формиране на професионални компетенции в областта на науката за данни са съществени онези компоненти на компетенциите (знания, умения, отношение и поведение), които се постигат в хода на аудиторното и практическото обучение по специалността.

Информационно-дидактическият компонент описва информационно-ресурсното осигуряване при отчитане на целите на системата в два основни подкомпонента:

- *информационно-ресурсен* – включва съдържанието, предвидено за усвояване от студентите в хода на обучението;
- *организационно-познавателен* – описва учебни задачи (или апарат) за организиране на усвояването на съдържанието от страна на студентите.

Уточняването на съвкупността от професионални компетенции, които изграждат професионалния профил на специалиста по данни, е задача, чието изпълнение е задължително в периода на планиране на обучението, базирано на компетенции. В хода на изследването на съществуващите образователни модели стана ясно, че по-голямата част от моделите, стремящи се към изграждане на компетенции, съдържат между 8 и 16 компетенции. Моделите включват дефиниции, често групирани в „кълъстери“, заедно с дескриптори на поведението [169].

Ключовият момент при изграждането на образователна методология е създаването на каталог от компетенции, съобразен с целите на

определената университетска програма. Показателите за поведение илюстрират как дадена компетентност се трансформира от абстрактно понятие в конкретни, ясни и осезаеми действия, които описват профила на успешния студент.

Каталогът от компетенции се състои от набор от *общи* и *специфични компетенции*:

- *Общите компетенции* са едни и същи за различните специалности в даден университет или в рамките на конкретно професионално направление.

- *Специфичните компетенции* засягат онези знания и умения, които са специфични за конкретния професионален профил.

Общите и специфичните компетенции, изисквани и очаквани от специалистите по данни, са подробно разисквани в раздел 3.1.3 на настоящата монография, а разработеният каталог от компетенции се намира в Приложение 3.

Изграждането на каталог от компетенции подпомага процеса на оценка на пропуските в знанията и уменията на студентите. От друга страна каталогът от компетенции ще уточни разминаванията между компетенциите, търсени от бизнес организациите и предлаганите такива в рамките на университетското образование. В своето изследване Nikolov et al. [169] предлагат този процес да бъде реализиран с помощта на матрица на пропуските (таблица 4.1.)

Таблица 4.1. Матрица на пропуските [169]

	Компетенция 1	Компетенция 2	...	Компетенция N
Дисциплина 1				
Дисциплина 2				
...				
Дисциплина N				

В редовете на матрицата са изброени всички курсове (дисциплини) от програмата, предвидени за съответната степен. В колоните са описани изискваните от специалиста професионални компетенции. Всички включени дисциплини, както и очакваните знания и умения от обучението в магистърската програма „Анализ и обработка на данни“, трябва да бъдат съпоставени с онези компетенции, които са значими и търсени от бизнес организациите. Това ще позволи привеждане на традиционните цели на обучението към нужните компетенции, а попълнената матрица може да се разглежда като основа за определяне на препоръки при проектирането на учебната програма.

След като приключи процесът по установяване на съвкупността

от компетенции, които ще бъдат включени в учебната програма, може да се премине към нейното проектиране. Тук трябва да се отчете фактът, че подходът, основан на компетенции, се фокусира основно върху резултата от обучението, който не се изразява в количеството научена информация от студента, а в неговата способност да действа адекватно в различни ситуации в професионален контекст. Завършващите специалисти не се нуждаят само от основни теоретични знания, но и от набор от компетенции, което им позволява прилагат и свързват мултидисциплинарни знания при нови и непознати предизвикателства. Това налага включените в програмата за обучение дисциплини да формират конкретни компетенции, които може да се проследят.

Оценката на резултатите от обучението представлява централна част в модела на обучение, основан на компетенции. Отчитането на постигнатите резултати има пряка връзка както с процеса на обучение, така и с компонента за контрол и обратна връзка от обучението. Всяка отделна дисциплина трябва да завършва с две оценки:

- *обобщаваща оценка за успеваемостта на студентите*, която обобщава постигнатите от тях компетенции;
- *оценка на проведения курс*, която позволява осъществяването на корекция на съдържанието на курса.

Решаващият елемент в модела, основан на компетенции, е използването на критерии, които позволяват поставянето на оценка. С оглед на спецификите на областта и целите на обучението в настоящия модел, за оценка на компетенциите ще бъде прилаган подходът за директно оценяване. При този подход се очаква студентите да докажат, че са усвоили изисквания набор от компетенции. В идеалния случай тази оценка трябва да бъде комбинация от три функции: функция за изпълнението, формираща и обобщаваща функция. Методите за оценяване, подходящи за обучението, са тестове, усни презентации и лабораторни доклади, анализи на текстове, данни или електронни портфейли.

За формирането на комплексна оценка на проведеното обучение е необходимо да бъдат осъществени и вътрешен одит и самоконтрол на обучението. Оценката предоставя на ръководителите на програмата обратна връзка, която позволява подобряването и развитието на учебния план и включените в него учебни програми. Във фазата на самоконтрол е необходимо да бъде осъществен и контрол на пропуските. Той ще позволи изготвянето на оценка на пропуските в обучението и ще търси разминаване между учебния план/програма и изискванията на бизнес организациите. За целта може да бъде из-

ползвана матрицата на пропуските, показана по-рано. Оценката ще позволи повишаване на качеството, ефективността и отчетността на предлаганото обучение, студентите ще получат достъп до необходимите за лесна реализация на пазара на труда компетенции.

Подготовката на учебен план на магистърска програма налага да се обърне внимание на това кои ще бъдат потенциалните кандидати за специалността и какви ще бъдат техните входящи умения. Знанието по тези въпроси ще позволи да бъде оптимизиран процесът по формиране на професионални компетенции за работа с данни. Този етап не е включен в общия модел на системата за формиране на професионални компетенции, но е част от процеса на проектиране на учебния план на магистърската програма.

4.1.3. Потенциални кандидати за магистърската програма

Изследването на големите данни, достъпни днес, позволява постигането на знания относно процесите, мотивацията, причините и следствията, а основното предизвикателство пред бизнес организациите е осигуряването на необходимия темп, ефективност и ефикасност на операциите. Технологиите и инструментите за съхранение и обработка на големи масиви от данни все още не позволяват техният анализ да се осъществява в отсъствие на човешкия фактор. Това изисква разбиране на човешките способности при търсенето на информация и знания и формулиране на ограниченията при работата с данни.

Работата с данни изисква специфичен набор от знания, умения и компетенции, които не са част от общия списък с компетенции. Това често поставя лицата, които четат и анализират данни, в позицията на медиатори между суровите данни и хората, нуждаещи се от техния анализ. Във връзка с това ние намираме допирни точки между информационния брокер и специалиста по данни.

Информационните брокери играят решаваща роля в медиацията между информационните ресурси и потребителите на информация. Този вид експертиза изисква специално обучение и професионално ориентирано образование [52]. Нарастващата сложност на външната среда налага осъществяването на скоростна обработка на информация и вземането на бързи и рационални решения. Инфоброкерите имат задачата да помагат на клиентите да се справят с предизвикателствата на информационната ера. Основната роля на информационните медиатори е да обслужват клиентите си, като представят в разбираема форма и вид същността на информацията, свързана с проблемите на дадения клиент [53].

Основните сходства между информационния брокер и специалист по данни са свързани със задачата на двата вида професионалисти да обслужват клиенти или част от дадена общност, като изследват наличните информационни ресурси. Те имат за цел да извлекат от информационния масив стойност, която е значима за крайния клиент. Знанията, генерирани вследствие на осъществените операции, подпомагат вземането на стратегически важни решения. На практика всяка от двете професии налага използването на широк набор от уменията, които ограничихме по-рано в четири категории (твърди, меки, аналитични и етични).

Интересен за нашето изследване се оказва въпросът дали е възможно професионалният профил на специалиста по данни, практикуващ в областта на науката за данните (Data Science), да бъде логично продължение на профила на инфоброкера, практикуващ в областта на науката за информирането (Informing Science). Зададен по друг начин, въпросът звучи така: Възможно ли е бакалавърска степен в областта на науката за информирането да служи като основа за надграждане на уменията в магистърска степен в областта на Науката за данните?

За да отговорим на този въпрос, е необходимо да определим изходящите умения на завършващите студенти от ОКС „Бакалавър“, специалност „Информационно брокерство“. Данни за определяне на изходящите умения можем да съберем чрез преглед на учебния план на специалността, предложен от Университета по библиотекознание и информационни технологии. Анализът ще бъде осъществен в три стъпки:

- преглед на дисциплините, заложен в учебния план, и разпределение на дисциплините в една от четирите категории компетенции;
- анализ на разпределението на дисциплините по категории;
- съпоставка на разпределението с очаквания набор от компетенции в магистърската степен.

Разпределението на всички дисциплини по групи показва, че най-голям дял заемат дисциплини, изграждащи твърди умения (60%), следвани от дисциплини, формиращи меки умения (31%). По-малък дял в учебния план на инфоброкерите заемат аналитичните умения (6%), а най-малко присъствие имат етичните умения (3%). Разбивка по категории на дисциплините, изграждащи учебния план на специалността „Анализ и управление на данни“, е представена в Таблица 4.2.

Таблица 4.2. Дисциплини, формиращи учебния план на магистърска програма „Анализ и управление на данни“

Твърди (технически) умения			Меки (нетехнически) умения		Аналитични умения	Етични умения
Статистика и математика	Програмен език	Основи на ИТ	Процеси на информиране	Бизнес отношения и комуникации	Анализ	Етика
Дискретна математика	Основи на програмирането	Компютърни архитектури	Въведение в информационното брокерство	Английски език	Математически анализ	Правни аспекти на ИТ
Линейна алгебра и аналитична геометрия	Обектно ориентирано програмиране	Мобилен технологии	Информационно брокерство	Основи на електронния бизнес	Системен анализ	
Теоретични основи на информатиката	Визуални среди за програмиране	Уебтехнологии	Информационен мениджмънт	Основи на електронното правителство		
Теория на вероятностите и математическа статистика	SQL	Операционни системи	Организация и управление на знанията	Системи и компетенции за ИТ кариерно развитие		
Математическо моделиране	Структури от данни и алгоритми	Бази от данни	Процеси и системи за информиране			
SAS	ASP I част	Компютърни мрежи и комуникации	Теория и практика на консултирането			
		Информационни системи				

Базирайки се на каталога от компетенции, представен в *Приложение 3*, и на анализа на нужните умения, изграждащи профила на специалиста по данни, можем да твърдим, че за да бъде надграден профилът на информационния брокер в специалист по данни, е необходимо да бъде осигурено обучение с акцент върху някои съществени за профила на специалиста по данни умения, сред които:

- твърди умения – обучение по дисциплини, обхващащи големи данни, статистика, машинно обучение и облачни технологии;

- аналитични умения – дисциплини, свързани с процесите обработка, извличане, интерпретация, визуализация и анализ на данни.

Посочените дисциплини не са застъпени в обучението на инфоброкерите в бакалавърската програма „Информационно брокерство“, но са съществен елемент от профила на специалиста по данни.

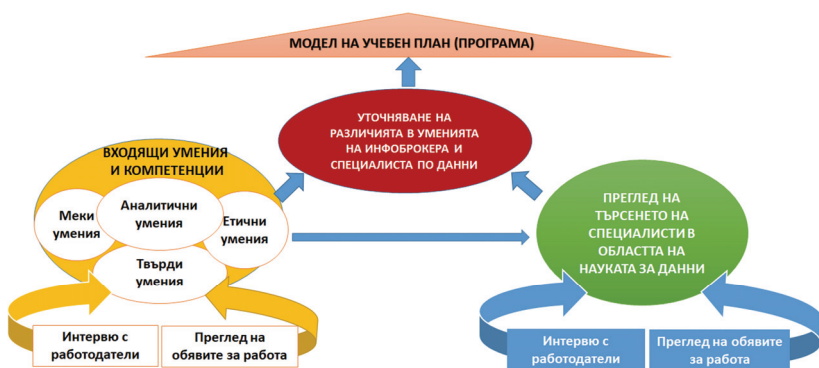
Информационните брокери са обучени да обработват информация, но не са готови да работят с данни, а още повече с големи данни. В хода на своята бакалавърска програма те изучават редица дисциплини, изграждащи техните меки умения, но дисциплините с акцент върху твърдите умения не отговарят на профила на специалистите, занимаващи се с данни. За да могат да се реализират като специалисти по данни, е необходимо студентите да преминат обучение, адаптирано към техните входящи знания и умения. Трябва да се изпълнят няколко основни стъпки:

- анализ на входящите умения на потенциалните кандидати за специалността „Анализ и управление на данни“;

- анализ на набора от умения, позволяващи работа с данни, изисквани от бизнес организациите;

- определяне на списък от задължителни умения в профила на специалиста по данни, който информационният брокер не притежава;

- създаване на учебна програма и обучение, насочени към изграждането на компетенции в областта на науката за данните (Фиг. 4.4).



Фиг. 4.4. Карта за надграждане на компетенциите от науката за информиранието към науката за данните

Информацията, включена в Таблица 4.2, ни дава основание да смятаме, че специалистите, доставящи информация (т.нар. информационни брокери) покриват част от търсените умения и компетентности от профила на съвременния специалист по данни.

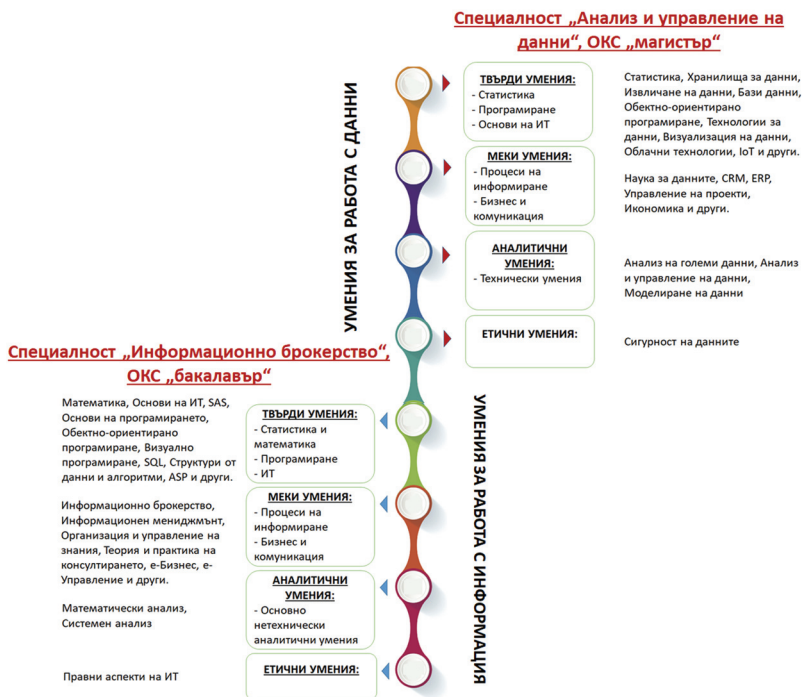
За да отговори на предварително зададените от нас условия, учебният план на магистърската програма „Анализ и управление на данни“ трябва да бъде изграден така, че да допълва знанията и уменията на студентите, завършващи специалност „Информационно брокерство“. Това налага магистърската програма в областта на науката за данните да бъде разглеждана като логично продължение на знанията и уменията на студентите, завършващи бакалавърската програма в областта на науката за информизиране. На практика, магистърският курс трябва да надгражда вече овладени в бакалавърската степен умения, без да дублира изучавани дисциплини.

На базата на резултатите от предварителните анализи, представени дотук, е необходимо да бъде разработен учебен план на магистърска програма „Анализ и управление на данни“, който да попълва установените пропуски в компетенциите. Списък на дисциплините, които ще позволят това надграждане на професионалния профил на информационния брокер в специалист по данни, е представен в следващата таблица.

Таблица 4.3. Дисциплини, надграждащи професионалния профил на информационния брокер

Твърди умения			Меки умения	Аналитични умения	Етични умения
Статистика и математика	Програмни езици	Основи на ИТ	Информизиране, Бизнес и комуникации	Анализ	Етика
Статистика	Хранилища на данни/Data Warehousing	Визуализация на данни. Ер-гономия	Въведение в науката за данни	Анализ на големи данни	Разкриване на заплахи и измами
Теория на вероятностите и статистика*	Извличане на данни/Data Mining	Архитектура на приложения	Поведенческа икономика	Разработка и управление на изискванията	
Математически анализ*	Бази данни*	Въведение в облачните технологии	Управление на отношенията с клиенти (CRM)	Анализ и управление на данни в ДА	
	Обектно ориентирано програмиране*	Технологии за данни	ERP системи	Моделиране и симулиране в графична среда	
		Интернет на нещата	Създаване и управление на проекти		

Както може да се види, акцентът е поставен върху изграждането на твърди умения (55%) и в по-малка степен – на меки (23%) и на аналитични (18%) умения. Програмата включва всички онези дисциплини, които анализът, осъществен от нас по-рано, определи като необходим минимум за изграждането на профила на специалиста по данни. Широко застъпени са дисциплини, засягащи процесите на съхраняване, извличане, управление и анализ на данни.



Фиг. 4.5. Надграждане на уменията на инфоброкера

В обобщение можем да кажем, че двете специалности – „Информационно брокерство“ и „Анализ и управление на данни“, може да се разглеждат като трансдисциплинни специалности, които взаимно се допълват. В своето единство те осигуряват уменията, изграждащи профила на съвременния специалист по данни.

В хода на изграждане на учебния план на магистърската програма бакалаврите инфоброкери ще бъдат разглеждани като част от групата на потенциалните кандидати за специалност „Анализ и управление

ние на данни“. Това ни дава основание да разглеждаме изходящите умения на информационните брокери като входящи умения на кандидатите за проектираната специалност.

4.2. ПРОЕКТИРАНЕ НА МАГИСТЪРСКА ПРОГРАМА „АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“

4.2.1. Проектиране на учебен план – предпоставки и нормативна уредба

Изграждането на компетенции в хода на обучението може да бъде реализирано чрез проектиране на цялостна учебна програма, основана на компетентностния подход, както и чрез индивидуален дизайн на всяка от включените дисциплини. При проектирането на учебния план на програмата е необходимо като първа стъпка да бъдат осъществени формулиране и групиране на очаквания набор от компетенции. Като следваща стъпка е нужно изготвянето на анализ на всяка компетенция с цел разделянето ѝ на различни нива на сложност.

В настоящия раздел ще бъде представен процесът на подготовка и изграждане на учебен план на магистърска програма по „Анализ и управление на данни“, адаптиран към нуждите на Университета по библиотекознание и информационни технологии и насочен към развитието на аналитични компетенции в областта на науката за данните.

За да се премине към проектирането на магистърската програма, е необходимо да се отговори на някои съществени въпроси, засягащи процеса на обучение. Най-често тези въпроси се свързват с дейности като (фиг. 4.6):

- *определяне на целта на обучението* – дефиниране на компетенциите, които трябва да бъдат овладени в хода на обучението;
- *дефиниране на учебните дейности* – формулиране на методите и средствата, които ще служат за изграждане на компетенциите;
- *планиране на учебния процес* – определяне на формата на обучението, продължителността и броя кредити, формите на контрол на придобитите знания/умения и др.;
- *отчитане на резултатите от обучението* – анализ на овладените от обучаемите компетенции в хода на обучението.



Фиг. 4.6. Процес на проектиране на учебна програма

Магистърската програма е предназначена за студенти, завършили ОКС „Бакалавър“ в УниБИТ и други висши училища, както и за работещи в различни организации на позиции, изискващи работа с данни. **Основната цел** на програмата се свързва с осигуряването на знания, компетентности и умения за:

- организиране на големи информационни масиви в условията на разнообразни ИКТ платформи;
- организиране на събирането и осигуряването на качеството на данни;
- извличане на полезни знания от данните със статистически и други аналитични методи;
- оценяване на чувствителността и доверието към резултатите от анализа на данните;
- проектиране, разработване, изграждане и използване на компютърни приложения, предназначени за анализ и управление на данни в различни предметни области.

Учебните дейности и процесът на обучение са съобразени с предварително утвърдени учебен график и учебна документация. Учебният план, като краен продукт, трябва да осигурява на обучаемите възможност да постигнат определени учебни цели чрез извършването на определени учебни дейности в определена последователност в контекста на определена учебна среда [170].

Учебният план на магистърската програма е проектиран за нуждите на акредитиран от Националната агенция по оценка и акредитация (НАОА) университет в България (с държавно финансиране). Това налага да се проучат действащото законодателство и изискванията на институциите по отношение на тази степен на обучение.

Обучението във висше училище следва нормативна рамка, наложена от Министерството на образованието и науката (МОН) чрез Закона за висшето образование [7]. Законът урежда устройството, функциите, управлението и финансирането на висшето образование в Република България.

За целите на нашето изследване законът е интересен по линия на наложени ограничения върху:

- очаквания набор от документи за стартиране на магистърска програма;

- очаквания брой кредити;
- очаквания брой часове;
- учебното съдържание и продължителността на обучението;
- формите за проверка и оценка на знанията и уменията на студентите;
- завършването на студентите.

В следващата таблица са приложени извадки от Закона по отношение на посочените по-горе показатели.

Таблица 4.4. Извадки от ЗВО, свързани с обучението в магистърска степен

Показател	Изисквания на ЗВО
Образователна степен	(3) (Нова – ДВ, бр. 60 от 1999 г.) Обучението в степента „магистър“ осигурява задълбочена фундаментална подготовка, съчетана с профилиране в определена специалност.
Учебно съдържание	Чл. 40. (1) Учебното съдържание по всяка дисциплина се организира в относително обособени части. (2) Всяка част следва да включва минимум 15 академични часа аудиторна заетост. Академичният час е с продължителност 45 минути. Чл. 41. (1) Във висшите училища обучението се извършва по задължителни, избираеми и факултативни дисциплини. (2) Висшите училища осигуряват за всеки студент възможности за избор на учебни дисциплини в рамките на утвърдения учебен план.
Учебна документация	(2) (Изм. – ДВ, бр. 60 от 1999 г., предишен текст на чл. 39, бр. 48 от 2004 г.) Учебният процес във висшите училища се провежда по учебна документация за всяка специалност, която обхваща квалификационни характеристики по степени, учебен план, учебни програми на изучаваните дисциплини и ежегоден график на учебния процес.
Образователни кредити	Чл. 44а. (Нов – ДВ, бр. 48 от 2004 г., в сила от учебната 2004/2005 г.) (1) За оценка на придобитите по време на обучението знания и умения на студентите висшите училища осигуряват система за натрупване и трансфер на кредити. (2) Целта на системата е да осигури на студентите възможност за изборност на дисциплини, самостоятелна работа и мобилност на основата на взаимно признаване на отделни периоди на обучение, които

Показател	Изисквания на ЗВО
	<p>отговарят на част от учебния план на специалността и образователно-квалификационната степен.</p> <p>(3) Кредитите са оценка на задълженията, които студентите трябва да изпълнят по дадена дисциплина или модул като част от общите задължения, необходими за успешно завършване на една учебна година или семестър.</p> <p>(4) По всяка дисциплина кредитите се формират от хорариум от лекции и от практически и семинарни упражнения, самостоятелна работа (в лаборатории, библиотеки, курсови, домашни работи и други), положени изпити и други форми на оценяване, определени от висшето училище.</p> <p>(6) Системата по ал. 1 се основава на 60 кредита за учебна година, или 30 кредита за семестър, разпределени по учебни дисциплини и/или модули съгласно учебния план на специалността и образователно-квалификационната степен или на друг съвместим с тази система брой кредити.</p>
Форми за проверка и за оценка на знанията и уменията на студентите	<p>(2) (Изм. – ДВ, бр. 41 от 2007 г.)</p> <p>Формите за проверка и за оценка на знанията и уменията на студентите се определят в учебните планове и програми. Основна форма на оценяване на знанията са изпитите, които са писмени, освен когато спецификата на учебната дисциплина и/или модул изисква друго.</p> <p>(3) Знанията и уменията на студентите се оценяват по шестобална система, която включва: отличен (6,00), много добър (5,00), добър (4,00), среден (3,00) и слаб (2,00).</p> <p>(4) За успешно положен се смята изпитът, оценен най-малко със среден (3,00).</p>
Завършване	<p>Чл. 45. (1) (Изм. – ДВ, бр. 41 от 2007 г.) Обучението по всяка специалност на образователно-квалификационна степен „бакалавър“ или „магистър“ завършва с държавен изпит или със защита на дипломна работа.</p>

Въз основа на извършения преглед на ЗВО и направените извадки от него може да бъдат изведени следните обобщения:

- магистърската програма трябва да предлага профилиране в специалността „Анализ и управление на данни“;
- всяка от дисциплините, включени в учебния план на специ-

алността, трябва да покрива минимум 15 академични часа аудиторна заетост;

- учебният план трябва да включва задължителни, избираеми и факултативни дисциплини;

- официалната учебна документация на магистърската програма трябва да включва: квалификационна характеристика, учебен план, учебни програми на изучаваните дисциплини и ежегоден график на учебния процес;

- кредитите, покривани чрез дисциплините, заложен в учебния план, не трябва да бъдат по-малко от 60 за учебна година, или 30 за семестър;

- основната форма на оценяване на знанията са изпити (писмени или практически), като изпитната форма трябва да бъде определена в учебната програма на всяка дисциплина.

Очакваният резултат от обучението по „Анализ и управление на данни“ се свързва с постигането на конкретни знания, умения и компетенции.

4.2.2. Учебна документация

Всеки учебен план е структуриран от конкретен набор от дисциплини, които може да се разглеждат като единици за обучение. В този аспект проектирането на магистърската програма може да се разглежда като моделиране на единици за обучение [168], а всяка от единиците – като фундамент за формиране на конкретна компетенция.

Опирайки се на ЗВО, е необходимо в хода на проектирането на учебния план по „Анализ и управление на данни“ за ОКС „Магистър“ да бъдат разработени последователно следните документи:

- квалификационна характеристика;
- учебен план;
- учебна програма за всяка от участващите в плана дисциплини.

За всяка учебна единица е необходимо да бъде осигурена специално проектирана учебна програма, основаваща се на специфичните методи и очаквани знания, умения и компетенции. Това е задача на ръководителя на съответната дисциплина, включена в учебния план на програмата (фиг. 4.7).

Задължение на ръководителите на съответните дисциплини, включени в обучението, е да подготвят учебна методология и учебни ресурси, необходими за нормалното протичане на учебния процес.



Фиг. 4.7. Проектиране на учебен план – основни документи и ресурси

А. Квалификационна характеристика на магистърска програма „Анализ и управление на данни“

Квалификационната характеристика на магистърската програма (Приложение 5) е изработена съобразно форма, възприета в Системата за управление на качеството на обучение (СУКО) в УниБИТ и включва следните раздели:

- Анотация;
- Знания;
- Компетенции;
- Умения;
- Перспективи за реализация.

Квалификационната характеристика, изградена за нуждите на обучението в магистърската програма, изисква постигането на: „специализирани теоретични познания по всички аспекти, свързани с изграждането на бизнес информационни системи, ориентирани към управление, базирано на данни; софтуерни приложения, създадени за анализ на данни; статистически и интелигентни методи, модели и техники за извличане на знания от данни, а също и техники за визуализация на резултатите, ориентирани към лесно и интуитивно възприемане на резултатите от анализа на данни“ (извадки от квалификационната характеристика, Приложение 5).

Основните компетенции, които студентите трябва да придобият в хода на своето обучение, са да:

- използват съвременни методи за анализ и управление на данни;

- разработват и реализират успешни стратегии в областта на организация на процесите по получаване, верификация, структуриране, съхраняване и управление на информационни ресурси;
- анализират данни и правят изводи въз основа на данни с помощта на компютърни технологии;
- използват различни платформи за развиване на нови софтуерни продукти и услуги в областта на анализа на данни;
- събират, анализират и интерпретират информация, необходима за изгълнителска, организационно-управленска, консултантска и научноизследователска дейност;
- формират информационно поведение и информационна култура в обществото; анализират и използват специални информационни масиви и др.

Квалификационната характеристика поставя рамка и на придобитите ключови умения:

- интегрално прилагане на придобитите знания за анализ и управление на данни чрез информационни технологии в работата;
- ефективна работа по анализ и планиране на развитието на информационните ресурси;
- решаване на проблеми в областта на управлението на данни;
- работа в екип, формулиране и вземане на самостоятелни решения;
- обучение през целия живот и ефективно споделяне на знания, технологии и добри практики.

В квалификационната характеристика на специалността следва да бъде отразено, че успешно завършилите магистри ще притежават задълбочени аналитични знания и умения и ще успеят успешно да се реализират във всякакви институции, нуждаещи се от знания, компетентности и умения за организиране, управление и анализ на данни. Потенциалните работодатели на тези специалисти са институции, банки, застрахователни дружества, телекоми и др.

Разработената квалификационна характеристика съпровожда учебния план на магистърската програма и може да бъде открита в Приложение 5 на настоящата монография.

Б. Учебен план

„Анализ и управление на данни“ е магистърска програма в област 4. „Приложни науки, математика и информатика“ на висшето образование, професионално направление 4.6. „Информатика и ком-

пютърни науки“, и е създадена за целите на обучението на студенти във Факултет „Информационни науки“ към УниБИТ. Въпреки принадлежността на новоизгражданата магистърска програма към технически ориентирано направление, ние смятаме, че тя не бива да ограничава желаещите по отношение на входящите им умения. Още повече че в Глава II показахме, че уменията за анализ на данни не бива да бъдат разглеждани единствено като специфика на технически ориентирани професионални профили. Тези компетенции не се ограничават до определена дисциплина, специалност или професия. Днес те са важно изискване за всеки бранш и всяка професия.

Това ни даде основание да предложим не един, а два учебни плана: първият, адаптиран за нуждите на бакалаври, завършили същото професионално направление (*Приложение 4.1*), а вторият – за входящите умения на бакалаври, идващи от други професионални направления (*Приложение 4.2*).

Учебният модел на магистърска програма, представен като краен продукт от осъществената дотук изследователска работа, съответства на разбирането на авторите относно съдържанието на програма, която има за цел да обучи специалисти с различен опит и експертиза. Професионалната квалификация на завършващите специалността е „Анализатор и мениджър на данни и информация“.

Учебният план, подобно на всеки друг официален документ, е изграден съобразно наложена от СУКО на УниБИТ форма за изготвяне на учебен план и включва информация за:

- *брой дисциплини по вид* – задължителни, избираеми и факултативни;
- *кредити* – съгласно поставените изисквания за трансфер на кредити 1 ECTS се равнява на 25 учебни часа;
- *хорариум на дисциплините* – определяне на аудиторната и извънаудиторната планирана заетост на студентите:
 - *аудиторна заетост* в семестъра: АО – общ брой; от тях: Л – лекции, У – упражнения, П – учебна и специализираща практика;
 - *извънаудиторната заетост* включва: консултации с преподавател; самостоятелни упражнения, подготовка в библиотеки, посещение на публична лекция и др.; самостоятелна работа (написване на публикация, реферат, курсова работа, друг вид разработка); самостоятелна подготовка;
- *форма на контрол и брой текущи изпити* – оценката се формира от: текущ контрол (тестове, колоквиуми, домашна работа и др.); защита на проект; защита на магистърска теза.

Учебният план предвижда обучение в три семестъра. Разпределението на учебните дисциплини според техния вид (задължителни, избираеми и факултативни) е предложено в следващата таблица.

Таблица 4.5. Разпределение на дисциплините по вид

	Задължителни дисциплини (бр.)	Избираеми дисциплини (бр.)	Факултативни дисциплини (бр.)	Общо дисциплини, осигуряващи необходимите кредити (З + И + Ф)
План за студенти, завършили ОКС „Бакалавър“ в друго професионално направление	12	10 (* От тях се избират 3 или 4.)	1	16
План за студенти завършили ОКС „Бакалавър“ в същото професионално направление	8	10 (* От тях се избират 3 или 4.)	1	12

Предварителните изисквания към кандидатите за обучение в магистърската програма се свързват с уеднаквяването на фундаменталната основа. Дисциплините, наложени като част от предварителните изисквания към специалността, са:

- Математически анализ;
- Теория на вероятностите и статистика;
- Дискретна математика;
- Обектно ориентирано програмиране;
- Структури от данни и алгоритми;
- Релационни бази данни и SQL.

Дисциплините, заложи за обучение в първия семестър, осигуряват общо 900 учебни часа в задочна форма на обучение. От тях 160 учебни часа са за лекции, 120 – за упражнения, а 620 – за извънаудиторна заетост. Разпределението на дисциплините от първия семестър е предложено в таблица 4.6.

Разпределението на дисциплините, включени в плана за обучение в рамките на втория семестър, е предложено в таблица 4.7.

Таблица 4.6. Дисциплини, включени в първия семестър на обучение

№	Дисциплина	Семес- тър	Кре- дита ECTS	Общо	Аудиторна заестост				Извън- аудиторна зае- гост	Форма на изпит- ване
					АО	Л	У	П		
1-ВИ СЕМЕСТЪР										
1.	Въведение в науката за данните	1	5	125	35	20	15	0	90	И
2.	Статистика	1	5	125	35	20	15	0	90	И
3.	Хранилища на данни (Data Warehousing)	1	5	125	35	20	15	0	90	И
4.	Извличане на данни (Data Mining)	1	5	125	35	20	15	0	90	И
5.	Визуализация на данни. Ергономия	1	5	125	35	20	15	0	90	И
6.	Математически анализ*	1	3	75	35	20	15	0	40	И
7.	Теория на вероятностите и статистика*	1	3	75	35	20	15	0	40	И
8.	Избираема дисциплина	1	5	125	35	20	15	0	90	И
Общо за 1-ви семестър			36	900	280	160	120	0	620	8И

Таблица 4.7. Дисциплини, включени във втория семестър на обучение

№	Дисциплина	Семес- тър	Кредити ECTS	Общо	Аудиторна заестост				Извън- аудиторна заестост	Форма на изпитване
					АО	Л	У	П		
2-РИ СЕМЕСТЪР										
1	Анализ на големи данни	2	5	125	35	20	15	0	90	И
	Архитектура на приложение	2	5	125	35	20	15	0	90	И
	Въведение в облачните технологии	2	5	125	35	20	15	0	90	И
	Бази данни*	2	3	125	35	20	15	0	40	И
	Обектно ориентирано програмиране*	2	3	75	35	20	15	0	40	И
	Избираема дисциплина	2	5	75	35	20	15	0	90	И
	Избираема дисциплина	2	5	125	35	20	15	0	90	И
	Избираема или факултативна дисциплина	2	5	125	35	20	15	0	90	И/Т
Общо за 2-ри семестър			36	900	280	160	120	0	620	7И + И/Т

В третия семестър на учебния план са включени изпълнение на научноизследователски проект и разработка на магистърска теза.

Таблица 4.8. Дисциплини, включени в третия семестър на обучение

№	Дисциплина	Семестър ECTS	Кре- дита	Общо	Аудиторна заетост				Извънау- диторна Заетост	Форма на из- питване
3-ТИ СЕМЕСТЪР										
	Научноизследователски проект	III	15	375	0	0	0	0	375	Зпр
	Магистърска теза	III	15	375	0	0	0	0	375	Змт
	Общо за 3-ти семестър		30	750	0	0	0	0	750	Зпр + Змт

За всяка дисциплина от учебния план се създава учебна програма, чиято форма се налага отново от действащата СУКО на УниБИТ.

В. Учебни програми

Всяка учебна програма включва описателна информация, като наименование, код, тип и равнище на курса, в коя година от обучението се появява и колко кредита осигурява. Изисква се също информация за ръководителите на курса на обучение, както и подробно разписан реферат на дисциплината, усвоени знания, умения, компетенции, предварителни изисквания и съдържание на курса. Съгласно вътрешен правилник на УниБИТ и ЗВО учебната програма се изработва от ръководителя на съответната дисциплина, който следи за актуалността и обхвата на съдържанието.

Г. График

Учебните планове и програми се утвърждават веднъж годишно от катедрен, факултетен и академичен съвет при строго определен график. Обучението на студенти по плановете и програмите се осъществява по издаден от Учебен отдел график, който е утвърден от академичен съвет и е публично оповестен чрез сайта на университета.

4.2.3. Управление на риска при обучението по анализ на големи данни

Управлението на риска обикновено е част от обучението на студентите, занимаващи се с управление на проекти. Такова обучение често включва преглед на специфични рискове в областта. Например специално внимание в ИТ проектите се отделя на проучване на риска, свързан с наличие на грешки в софтуера. Обучението на студентите в този случай е насочено към това как да съумеят да организират процеса на разработване на софтуера по начин, който да идентифицира и отстранява грешки.

В областта на науката за данните рискови са събития, които водят до погрешни или заблуждаващи решения. Такива събития може да възникнат като следствие от погрешно тълкуване на резултатите, получени от инструментите за анализ на данни. Като цяло рисковете, свързани с науката за данните, може да се разглеждат като резултат, зависещ от човешката готовност и експертиза. Често потребителят разглежда технологичната обработка на данни като „черна кутия“, без да влага ясно разбиране за това как даденото компютърно приложение произвежда резултатите, нито дори колко чувствително е то към данните (особено за нарушаване на математически условия по отношение на свойствата на данните за постигане на надеждни резултати).

При изграждането на учебен план на магистърска програма, чийто фокус е управлението на данни, е изключително важно да бъдат обхванати теми, изучаващи риска. В обучението по управление на риска в областта на науката за данните е нужно да се обхванат онези области на потенциални опасности, които не са обхванати от други дисциплини и които представляват специфичен риск за успеха при изучаването и анализа на големи данни. В тази посока успяхме да идентифицираме три области на опасност (риск):

- готовност на потребителя;
- качество на данните;
- чувствителност на аналитичните инструменти.

Готовността на потребителите определяме като ниво на осведоменост и готовност. Произходът на опасностите в тази категория се свързва с понятието „информационна асиметрия“. Специалистите по данни и крайните потребители често използват различни жаргони и термини. Неправилното подаване на информация може да доведе до погрешно тълкуване и подвеждане на получателя на съобщението, а то от своя страна да предизвика погрешни решения, водещи до нежелани загуби. Липсващата оценка за осведомеността и готовността на потребителя може да доведе до изграждане на интерфейси, които не вдъхват доверие, и в резултат на това потребителите не вземат решения, основани на данните.

Следващата опасност се свързва с **качеството на данните**. За да се приложи дадената аналитична техника, специалистът по данни трябва да „почисти“ данните. Често това води до загуба на значителна информация. В ерата на големите данни проверката на качеството на данните, оценката на свойствата им и прилагането на аналитични техники изискват специфични умения. Част от тези умения са: *проследяване на произхода на данните; оценка на достоверността на данните; точност; контекст на създаване и записване; фактори, влияещи върху данните* и много други. В някои случаи оценката на свойства като *прецизност, повторяемост и възпроизводимост* изисква знания за

използваното оборудване, компетенциите на хората, създали тези данни, и т.н. Невъзможността да се провери качеството на данните, създава риск от получаване на заблуждаващ резултат.

Рисковете при **обработката на данни** са предимно резултат от проучване на аналитични приложения, използвани като „черни кутии“. Най-често процесът на обработка на данните ни изправя пред процес, включващ следните задачи:

- осигуряване на данни;
- наличие на данни;
- осигуряване на софтуер за работа с тези данни;
- импортиране на данните в софтуера;
- получаване на резултати от данните чрез използване на софтуера.

Липсата на компетенции води до приемане на неподходящи аналитични техники, прилагане на техниката по грешен начин или приемане на резултатите, без да бъде отчетен ефектът на чувствителност.

Таблица 4.9. представя нашата визия за обучението по управление на риска в магистърската програма „Анализ и управление на данни“.

Таблица 4.9. Управление на риска в учебния план на специалност „Анализ и управление на данни“

Критерий	Риск	Загуби	Насоки/акцент в курса	Дисциплини, преодоляващи риска
Готовност на потребителите	Информационна асиметрия	Грешно тълкуване и заблуждение; неправилни решения и загуби	Как потребителите възприемат данните и контекста; какви са проблемите, с които се сблъскват те, и какви са приемливите решения	„Анализ на големи данни“; „Визуализация на данни. Ергономия“
Качество на данните	Загуба на значителна информация	Получаване на подвеждащ резултат чрез прилагане на аналитична техника за обработка на данни	Проследяване на произхода на данните, оценка на достоверността, точността, контекста на създаване и записване на данните, както и на факторите, влияещи върху данните	„Бази данни“; „Хранилища на данни“
Чувствителност на аналитичните инструменти	Липса на компетенции	Приемане на неподходящи аналитични техники, прилагане на техниката по грешен начин или приемане на резултатите, без да бъде отчетен ефектът на чувствителност	Коя техника е най-подходяща за конкретния случай, как да я приложим правилно при отчитане на целите, домейна, контекста и свойствата на данните; как да бъдат представени резултатите на потребителите, като се има предвид тяхната готовност и осъзнатост	Всички дисциплини, включени в програмата

Дисциплината „Анализ на големи данни“, от една страна, включва обучение на студентите, което е свързано с оценка на степента на зрялост на стопанския субект (клиента), а от друга страна, засяга и оценява готовността на конкретните потребители. Дисциплината „Визуализация на данни. Ергономия“ обучава студентите да прилагат различни техники, приложими за потребители от различен тип (от много наивни до силно недоверчиви). Изграждането на умения за оценка на компетенциите на потребителите е една от целите на тази дисциплина. Студентите трябва да насочат вниманието си и към разбиране за това как потребителите възприемат данните, контекста на данните; какви са проблемите, с които се сблъскват, и какви са приемливите за тях решения.

Рискът, свързан с качеството на данните, се обсъжда в дисциплините „Бази данни“ и „Хранилища на данни“. Изучават се няколко случая на нарушения на качеството на данните, които водят до погрешни решения и значителни загуби. Разглеждат се различни случаи, като наличието на редови данни, които не отговарят на изискванията за използване на дадени техники (например данни, които обикновено не се разпространяват при прилагане на статистически данни, силно чувствителни към това свойство на данните), както и ефекта, причинен от почитване на данните чрез премахване на непълни или липсващи записи. В учебните програми на разглежданите дисциплини са включени специфични техники за намаляване на тези рискове.

Цялостната програма по „Анализ и управление на данни“ е насочена към формирането на знания, умения и компетенции, позволяващи на студентите да:

- посочат най-подходящата техника за анализ на данни; разберат как да приложат тази техника по подходящ начин, като отчетат целите, областта, контекста и свойствата на данните;
- изберат подходящ начин за представяне на резултатите на потребителите, като имат предвид тяхната готовност и осведоменост.

Рисковете обикновено са последният компонент от обучението. Може да твърдим, че от една страна, анализът на данните е редовен оперативен процес, а от друга страна, всеки конкретен анализ на данни е специфичен, различен и изложен на много опасности. Рисковете не са пряко свързани с разработването на приложения, а с това как се анализират данните и как се възприемат и използват резултатите. Ето защо отчитането на потенциалните рискове е предизвикателство пред проектирането на учебна програма по анализ на данни.

4.3. КУРС ПО „ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ“

4.3.1. Проектиране на учебна програма на дисциплината

Обучението на студентите за работа с големи данни е сложна задача. Необходимите компетенции са доста разнообразни, а голяма част от необходимите знания и умения се изграждат в рамките на обучението в различни дисциплини. Обучението формира умения за: ефективна работа с технологиите; прилагане на сложни математически и статистически методи; разбиране и проучване на различни организационни техники; притежаване и прилагане на дълбоки познания в областта на науката за данните. Последно, но не по важност, е и задълбоченото разбиране на начина, по който потребителите възприемат информацията и намират ефективен начин за визуализиране на резултатите.

При разработването на учебната програма на дисциплината „Въведение в науката за данните“ трябва да се отчетат два факта:

- науката за данните е широка област, която не е обхваната изцяло от конкретна специфична „класическа“ образователна област;
- тенденциите сред студентите показват, че младото поколение се оттегля от изучаването на теми, свързани с анализ на данни, каквито са математиката и статистиката.

Учебната програма за науката за данните е изградена върху трите основни категории компетенции, определени по-рано:

- възможност за извличане на полезни данни от огромни и разнообразни хранилища, включително публични и частни, както и от добре и слабо структурирани източници;
- способност за осъществяване на проверка на получените данни и оценка на тяхното качество;
- способност за интерпретиране (картографиране) на получените данни според контекста (проблема) и според приложни подходящи аналитични техники за извличане на полезни модели, връзки или просто за повишаване на разбирането относно обстоятелствата, свързани с проблема.

Възможността да се извърши цялата обработка на данни чрез ИТ по един високоефективен и ефикасен начин, представлява съществена част от учебната програма.

Критичната компетентност е способността да се разберат свойствата на достъпните и получени данни. От своя страна разбирането на свойствата на данните включва възможността да се отговори на следните въпроси:

- Какво представлява качеството на данните?
- Какви са съответните критерии за оценка на качеството на данните на даден източник на данни или проблемна област?
- Какви са факторите и обстоятелствата, които влияят върху събирането и представянето на данните по определен начин?
- Как ще се използват данните?
- Как да се измери дали наличието на данни и проучването на данните отговарят на горните критерии, за да се гарантират значими изводи?

Успехът в ерата на големите данни изисква също така:

- компетенции, които позволяват получаването на смислени, полезни резултати от данни, които не отговарят напълно на най-високите критерии за качество;
- компетенции, които позволяват осъществяването на рационални изводи при несигурност.

Съставянето на учебна програма изисква предварителна формулировка на преследваната цел и очакваните резултати от обучението. При отчитане на осъществената обзорно-аналитична работа, в настоящата разработка е представена обща формулировка на целта и очакваните резултати от обучението по дисциплината „Въведение в науката за данните“ (вж. Приложение б). Формулировката на **целта** на обучението се свързва с:

- получаване на теоретични знания – цялостна представа за състоянието, обхвата и ролята на науката за данните в съвременното общество;
- изграждане на практически умения – раглеждане на казуси, позволяващи определяне на задачите за решаване на проблем, изискващ компетентност от областта на науката за данните.

Дисциплината „Въведение в науката за данните“ изгражда рамката на магистърската програма „Анализ и управление на данни“ и очертава мястото на всяка отделна дисциплина като компонент от науката за данните. Дисциплината обхваща широк кръг от теми – от история на еволюцията на разбирането за извличане на знания от данни до ролята на конкретни научни области. В дисциплината са застъпени и теми, засягащи ползите за обществото и социалните измерения на науката за данните.

Очакваният резултат от дисциплината е изграждането на цялостен поглед върху предизвикателствата, технологиите, научните постижения и ползите за обществото от постигнатото ниво на науката за данните.

В **структурно отношение** учебната дисциплина е съставена от общо 10 теми. Разпределението на теоретичната и практическата работа според заложените часове в учебната програма е предложено в следващата таблица.

Таблица 4.10. Разпределение на темите, включени в дисциплината „Въведение в науката за данните“

№	Тема	Лекции	Упражнения
1.	Въведение в дисциплината	3	
2.	Еволюция на технологиите за извличане на полезни знания от данни	3	3
3.	Еволюция на компютърните технологии, свързани със събиране, съхраняване, достъп до и извличане на знания от данни	3	3
4.	Статистика – наука за анализ на данни	3	3
5.	Големи данни – предизвикателства и ползи	3	3
6.	Хранилища на данни (Data Warehousing) – задачи и приложимост	3	4
7.	Анализ на данни (Data Mining) – задачи и приложимост	3	4
8.	Машинно обучение	3	4
9.	Езици и технологии (R, Python, Minors)	3	3
10.	Социални аспекти на Data Science. Филантропия на данни	3	3

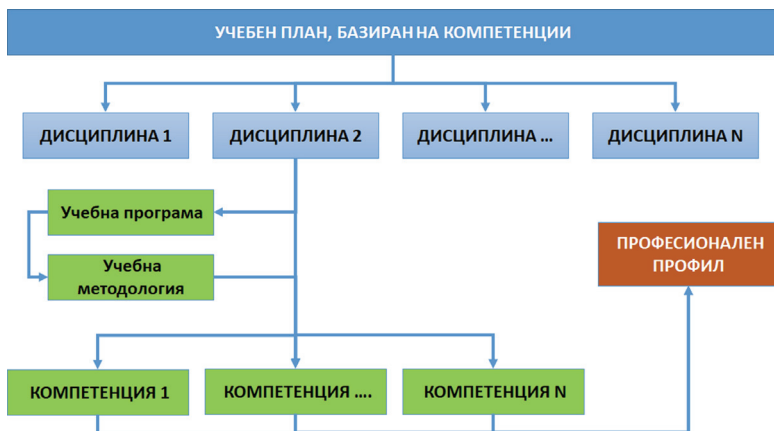
Часовете, заложи за изпълнение на учебно-преподавателската дейност по учебната програма, са с хорариум от общо 60 учебни часа, от които 30 учебни часа са за лекции и 30 – за упражнения. По правило, в учебната програма се поставя рамката на използваните методи и критерии на оценяване. Изпитът по „Въведение в науката за информиране“ включва подготовка на практически проект и изпитен тест.

Преподаването в областта на науката за данните е едно от най-големите предизвикателства днес. Причините за трудностите са от комплексен характер, но в същината си те се базират на предизвикателствата, стоящи пред големите данни, инструментите за техния анализ и готовността на специалистите да изучават данните. На свой ред хорариумът, наложен от учебната програма, е твърде малък, за да може да се разгърнат конкретни казуси с практическа значимост. Предизвикателства в общия случай се оказват също липсата на подходяща литература за преподаване и изучаване на науката за данните и липсата на добре подготвени преподаватели. Необходимо е ясно осмисляне на структурата на обучението и методите, прилагани от пре-

подавателите за формирането на компетенции у обучаемите. Наред със съставянето на учебна програма, при планирането на дисциплината трябва да бъде подготвена и методология на обучението.

4.3.2. Методология за обучение по „Въведение в науката за данните“

Създаването на **учебна методология** може да се разглежда като сценарий, по който следва да протича обучението. Препоръчва се [169] обхватът от учебни дейности в сценария да съответства на набора от поведенчески показатели за конкретна компетенция. Преподаването трябва да се прилага по такъв начин, че всяка дейност, която се извършва със студентите, да бъде насочена към една от компетенциите, които обучението трябва да формира. Мястото на учебната програма в обучението, основано на компетенции, може да бъде проследено на следващата фигура.



Фиг. 4.8. Мястото на учебната програма и учебната методология в обучение, основано на компетенции

Методологията на обучение, базирана на учебната програма по „Въведение в науката за данните“, разработена по теми, може да бъде открита в Приложение 7 на настоящата монография. В контекста на областта на науката за данните учебната програма, като крайна точка в картата, се отнася до изграждането на определени умения, които позволяват извличане на знания от наличните данни, и е насочено към обучението на студенти от Университета по библиотекознание и информационни технологии в България.

Обект и предмет на методологията

Обект на методологията е обучението по дисциплината „Въведение в науката за данните“, част от магистърската програма „Анализ и управление на данни“.

Предмет на методологията са логическите връзки и последователността на въвеждане на понятията в структурата на учебното съдържание по теми, свързани с големи данни и с възможностите за техния анализ, в учебната дисциплина „Въведение в науката за данните“. Обучението по дисциплината способства за формирането на основни знания за:

- обхвата и спецификите на науката за данните;
- историята от установяване на принципите на научно познание и появата на научни дисциплини, пряко свързани с извличане на знания от данни;
- технологиите за работа с данни от 1950 г. досега и очертаване на тенденциите в развитието им;
- основните задачи, които се решават с помощта на методите на статистиката;
- предизвикателствата, породени от масовото навлизане на компютърните технологии в живота, свързани с регистриране на всички факти в компютърни бази данни, известни като Big Data Challenge;
- технологиите, структурата от данни и организацията на приложения, ориентирани към подпомагане на процеса на вземане на решения, известни като „бизнес интелгентност“, чието ядро са хранилицата на данни;
- основните категории задачи, които се решават с инструментариума на Data Mining;
- основните категории задачи, които се решават с инструментариума, влизащ в категорията Machine Learning;
- съвременните езици, използвани за анализ на данни, средите, изградени около тях, и технологиите за подпомагане на дейностите по анализ на данни.

Принципи, залегнали при разработката на методология

При разработката на методологията са използвани формулировките на принципите на дидактиката, изложени в разработките от Андреев [2] и Петров [12]:

- принцип на научност;

- принцип на нагледност;
- принцип на съзнателност;
- принцип на активност;
- принцип на достъпност.

Всеки принцип ще бъде представен самостоятелно, като за основа ще бъде приложен обзорът, осъществен от В. Дзивев [6].

Принцип на научност

Съгласно този принцип по време на обучението на студентите трябва да бъдат предоставяни строго научни знания, факти, идеи, изводи, закономерности посредством методи, близки до методите на науката, чиито основи се изучават. Към този принцип трябва да се добави и изискването за точно спазване на терминологията, така както тя е описана в научната и техническата документация. Когато съществува термин, който няма съответствие на български език с международната терминология, терминът или задължително се съпътства с оригиналния термин, или се използва само оригиналният термин. Яснотата и еднозначността на терминологията са задължително изискване на научната методология [6].

Принцип на нагледност (визуализиране)

Нагледността е основа за всяко успешно обучение. В по-съвременните изследвания се публикуват доказателства за ефективността на този метод. Доказано е също, че използването на повече от една медия повишава степента на усвояване на учебния материал [158].

Принцип на съзнателност

Този принцип се основава на интелектуални качества и готовност за съзнателно отношение към усвояването и разбирането на учебния материал. Трябва да се създават условия, при които студентът да е отговорен и да може да работи съзнателно в процеса на обучението [6].

Принцип на активност

Същността на принципа се състои в единството между мислене и действие. Разглежда се неделимото единство между вътрешна и външна активност, между емоционална и интелектуална активност. Принципът на активността изисква непрекъснато стимулиране от страна на преподавателя. Той създава предпоставки студентът да реагира на поставяните задачи и да се стреми да получава адекватни резултати. Този принцип е силно зависим от професионализма на преподавателя.

Принцип на достъпност

Този принцип изисква обучението да се провежда в съответствие с възможностите на обучаемите. Учебният материал трябва да бъде с трудност, съответстваща на потенциала на обучаемите. Ако учебният материал е прекалено труден, се губи мотивация, а ако е лесен, се губи интерес и обучаемите не получават нови знания. Обучението като цяло трябва да е съобразено с особеностите на студентите, с равнището на тяхната подготовка и развитие. На базата на принципа на достъпност се формулират няколко правила за обучение [6]:

- от лесното към трудното;
- от известното към неизвестното;
- от простото към сложното;
- от близкото към далечното (в познавателен, а не в буквален смисъл);
- от конкретното към абстрактното;
- от фактите към изводите.

Принципът на достъпност изисква учебното съдържание да е съвременно и да е във връзка с настоящите достижения в разглежданата област.

Принцип на системност и последователност

Този принцип изисква знанията, уменията и навиците да се предлагат в определена последователност, като всеки елемент от учебния материал логически се свързва с другите.

Принцип на трайност на знанията, уменията и навиците

Този принцип изисква знанията и уменията да се усвояват трайно, за да благоприятстват по-нататъшното развитие на способностите. Знанията и уменията трябва да може да се прилагат в нови, различни ситуации. Този принцип способства за придобиване на умение за учене през целия живот. Това от своя страна спомага за поддържането на актуални знания, които са обвързани със съвременната действителност.

Принцип на индивидуален подход

Обучението трябва да е съобразено с индивидуалните особености на студентите и да подпомага тяхното развитие. Този принцип отхвърля подхода на стандартизация на преподавания материал. Пълната персонализация на преподаването е трудна, но трябва да се полагат усилия за свързване на „индивидуалност“ и „маса“ в образованието [6].

Цели, постигнати при изучаване на дисциплината „Въведение в науката за данните“

При разработката на методологията основните цели, които трябва да се постигнат, са:

- получаване на теоретични знания – цялостна представа за състоянието, обхвата и ролята на науката за данните в съвременното общество;
- изграждане на практически умения – разглеждане на казуси, позволяващи определяне на задачите за решаване на проблем, изискваща компетентност от областта на науката за данните.

Методи, използвани в обучението по „Въведение в науката за данните“

Методите, прилагани за **придобиване на нови знания**, може да бъдат [6]:

- *устно изложение на учебния материал от преподавателя* – тук може да бъдат съобщени факти, теория, да се формират определения и правила. В устното изложение се обясняват схеми и детайли от учебните помагала и материали, които се ползват. Детайлно се обясняват най-важните елементи по темата;

- *демонстрация* – едновременно се показват и обясняват елементите от преподавания материал;

- *беседа* – изразява се в задаване и отговаряне на въпроси. Форумират се въпроси, които да водят учениците към потвърждаване или отхвърляне на дадена теза, която се обсъжда. Трябва да се задават точни въпроси в правилна последователност. Въпросите не бива да съдържат непознати термини;

- *инструктаж* – при работа с техника, продукти, данни и др., когато съществува вероятност за неблагоприятни последствия, преди започването на практическа работа се посочват рисковете и начините за тяхното избягване;

- *самостоятелна работа с допълнителни информационни ресурси* – печатни и онлайн източници;

- *наблюдение* – студентите наблюдават конкретни действия.

Методите, прилагани за **формиране на понятия и умения**, може да бъдат [6]:

- *упражнение* – съзнателно и целенасочено повторение на дейности посредством различни видове задачи;

- *ситуационен метод*, свързан с вземане на решение при възникнали ситуации;

- *самостоятелна работа* – самостоятелно откриване на определение, правило, решаване на самостоятелна задача;
- *проучване* на готови модули и разкриване на характеристиката на използвани модули и детайли;
- *откриване на грешки* – умишлено заложи от преподавателя или възникнали случайно в процеса на работата;
- *експеримент* – проверка на функционалностите на различни модули и елементи в различни режими на работа. По този метод студентите имат възможност да откриват много нови знания;
- *дискусия* – възможно е провеждане на дискусия по теоретични проблеми, на които няма еднозначни решения, или по проблеми в науката, при които обучаемите сами достигат до решенията;
- *интерактивни методи* – решава се дадена задача, като студентите могат непрекъснато да задават въпроси или да споделят впечатления, свързани с резултатите от работата им. Така те едновременно получават знания, решават проблемни ситуации и взаимодействат помежду си, като се намират непрекъснато в активно състояние;
- *„мозъчна атака“* – това е вариант на интерактивния метод, при който се задава въпрос, с който студентите се мотивират да споделят своите идеи по някаква тема. Целта на този метод е не непременно да се получават само правилни отговори, а да се изследва целият спектър от мнения по въпроса. От друга страна, като подцел може да се разглежда активирането на аналитичното мислене и фантазията на студентите;
- *работа по проект* – студентите работят по конкретни казуси. Създават се условия за цялостно прилагане на учебния материал, както и възможност да се интегрира обучението по различните предмети, включени в него;
- *екипна работа* – формират се групи за реализиране на определена задача. В зависимост от сложността и обема на необходимата работа се преценява броят на участниците в една група. При краткосрочни проекти групите се подбират на случаен принцип, а при дългосрочни – по интереси, свързани с проекта;
- *ролеви методи* – студентите се поставят в различни роли: на потребители, администратори, анализатори и др. В тези си роли те трябва да решат дадена задача или да оценят качеството на определена система, която им е предоставена за работа;
- *визуализация* – връзките между модулите в потребителския интерфейс също трябва да се представят във визуална форма, способстваща тяхното възприятие;

- *мисловна карта* – описание на проектите като последователност от действия, роли на отделните участници в екипите и др.

Методите, предлагани за **проверка и оценка** на знанията, уменията и компетенциите, може да бъдат [6]:

- *текуща оценка* от вземане на участие по време на час;
- *оценка от решаване на задачи* по време на изпълнение на упражнения и участие в проекти;
- *оценка от участие* в дискусии, „мозъчни атаки“, експерименти и др.;
- *самооценка* – предоставяне на студентите на възможност да съставят предварително скала и регламент за оценка на дейностите и етапите при работа по проект и сами да оценят постигнатите резултати.

Организация на учебния процес по дисциплината „Въведение в науката за данните“

На база на гореизложените принципи, методи и правила и съгласно учебния план на дисциплината „Въведение в науката за данните“ е разработена методология за обучение, поместена в Таблица 4.11, структурирана по отделните теми.

Таблица 4.11. Организация и структура на методологията за изучаване на „Въведение в науката за данните“

№	Тема	Цел	Съдържание
1.	Въведение в дисциплината	Запознаване на студентите с целта, обхвата и съдържанието на магистърската програма и дисциплината „Въведение в науката за данните“	<p>Въведение:</p> <p><i>Тема 1.</i> Обем на данните (Volume). Взаимовръзка между обема на данните и информационните технологии. Исторически бележки</p> <p><i>Пример 1:</i> Холерит: преброяване на населението в САЩ през 1893 г.; изобретяване на перфокартите и сортировъчните машини</p> <p><i>Пример 2:</i> Емануел Голдберг, Карл Цайс – Йена, Rapid Selector</p> <p><i>Тема 2.</i> Динамика (Velocity). Развитие на комуникационните технологии: от 1800 г. до 21. век.</p> <p><i>Пример 3:</i> Комуникация, вземане на управленски решения в Източноиндийската компания</p> <p><i>Тема 3.</i> Разнообразие (Variety).</p>

№	Тема	Цел	Съдържание
			<p>Мултимедия, естествени езици <i>Обобщение: „3V“ модел на Gartner за големи данни</i></p> <p>2. Съдържание на магистърската програма по Data Science <i>Цел на програмата:</i> тематичен обхват; очаквани компетентности</p> <p>3. Съдържание на дисциплината: теми; теория; практика; изпитване; оценяване</p>
2.	Еволюция на технологиите за извличане на знания от данни	Запознаване на студентите с историята от установяването на принципите на научно познание и появата на научни дисциплини, пряко свързани с извличане на знания от данни	Преглед на основите на емпиричната и абстрактната наука – „източен“ и „западен“ път. Науката през вековете Обособяване на съвременните научни направления. Природни науки, математика, статистика, социални науки Процес на преобразуване на данни в знание: данни от наблюдения; данни от контролирани експерименти; построяване на модел; верификация на модела; използване на модела за прогнозиране Машинно обучение (Machine Learning) Статистическо машинно обучение Deep Learning Ръководено (Supervised) обучение Неръководено (Unsupervised) обучение Допълнителна тема: Управление на знанията. Теории за преобразуване на субективно (tacit) знание в записано (explicit)

№	Тема	Цел	Съдържание
3.	Еволюция на компютърните технологии, свързани със събиране, достъп до и извличане на данни	Преглед на технологиите, свързани с данни, от 1950 г. насам и очертаване на тенденциите в развитието им	Етапи в развитието на компютърните технологии Ранен етап – от машини, подпомагащи изчисленията, към машини за обработка на данни Данни, качество на данните, представяне на данните в компютърните системи Бази данни – основни идеи и изисквания. Йерархичен и мрежов модел Бази данни – реляционен модел и езикът SQL Интернет Ролята на проблема „Bug 2000“ за трансформация на базите данни 21. век: хетерогенни данни, облачни приложения, разпределени бази данни, социални мрежи
4.	Статистика – наука за анализ на данни	Преглед на основните задачи, които се решават с помощта на методите на статистиката	Науката статистика – кратък исторически преглед Основни задачи: построяване на модел, проверка на хипотези Анализ на данни от контролирани експерименти Анализ на данни от наблюдения Параметрични и непараметрични методи Основни направления: „Честотна“ статистика; Бейсова статистика Основни техники: регресия; класификация; прогнозиране
5.	Големи данни – предизвикателства и ползи	Запознаване с предизвикателствата, породени от масовото навлизане на компютърните технологии в живота, което е свързано с регистриране на всички факти в компютърни бази данни, известни като Big Data Challenge	Дефиниции на големи данни: „3V“, „5V“ и много „V“-та Технологии за обработка на големи данни Наука за данните: трансдисциплинарен характер компетентности

№	Тема	Цел	Съдържание
			<p>Големи данни и иновации</p> <p>Глобализация</p> <p>Предизвикателства</p> <p>Възможности</p> <p>Етапи в еволюцията на организациите по отношение на реализиране на възможностите при използване на големи данни</p>
6.	Хранилища на данни (Data Warehouses) – задачи и приложимост	Запознаване с технологиите, структурата от данни и организацията на приложения, ориентирани към подпомагане на процеса на вземане на решения, известни като „бизнес интелигентност“, чието ядро са хранилищата на данни	<p>Холистичен подход към решаване на бизнес проблеми</p> <p>Многомерни модели</p> <p>Кубове с данни</p> <p>Йерархии на дименсияте</p> <p>Процес ETL (Extract – Transform – Load)</p> <p>Изграждане на хранилища на данни:</p> <p>подход на Инмол;</p> <p>подход на Кембъл</p> <p>Визуализация</p> <p>Технологии: MS Power BI, Tableau, Click</p>
7.	Анализ на данни (Data Mining) – задачи и приложимост	Запознаване с основните категории задачи, които се решават с инструментариума на Data Mining	<p>Предварителна обработка: почистване на „мръсни“ данни;</p> <p>прогнозиране на липсващи данни;</p> <p>дискретизация;</p> <p>генерализация;</p> <p>редуциране на размерността и обема</p> <p>Характеризиране и сравняване</p> <p>Класификация и прогнозиране</p>
8.	Машинно обучение	Запознаване с основните категории задачи, които се решават с инструментариума, влизащ в категорията Machine Learning	<p>Статистическо машинно обучение (Statistical Machine Learning): построяване на модели, описващи изследваните обекти</p> <p>Deep Learning: обучение на „робот“ (компютърно приложение) да взема решения. Изкуствени невронни мрежи</p> <p>Контролирано машинно обучение (Supervised Machine</p>

№	Тема	Цел	Съдържание
			Learning) Неконтролирано машинно обучение (Unsupervised Machine Learning) Процес на обучение (Training and Testing Samples)
9.	Езици и технологии (R, Python, Data Miners)	Запознаване със съвременните езици, използвани за анализ на данни, със средите, изградени около тях, и с технологиите, които се развиват за подпомагане на дейностите по анализ на данни	Езикът R и обществото, поддържащо библиотеките с ресурси – програми за анализ на данни. RMiner Езикът Python, библиотеките Panda и Anaconda Rapid Miner, Knime, Orange и др.
10.	Социални аспекти на Data Science. Филантропия на данни		Data Science и организационната ефективност Значение на публично-частното партньорство Филантропията на данни като отговор на социалната отговорност към обществото на частните корпорации, притежаващи данни и допълнителни ресурси

РЕЗУЛТАТИ И ИЗВОДИ ПО ГЛАВА IV

В тази глава беше предложен модел на магистърска програма „Анализ и управление на данни“, насочена към изграждането на компетенции в областта на науката за данните. Изработването на учебната програма се базира на компетентностния модел и формирането на необходимите компетенции за бъдещите специалисти по данни.

В главата са постигнати следните по-съществени резултати:

1. Изградена е основана на компетенции методология за разработване на учебната програма.
2. Проектиран е общ модел на система за формиране на професионални компетенции в областта на науката за данните.
3. Представен е профилът на потенциалните кандидати за магистърската програма и са анализирани техните входящи умения.
4. Проектиран е учебен план на магистърска програма „Анализ и управление на данни“, който отчита входящите умения на кандидатите и е насочен към формиране на професионални компетенции в областта на анализа на големи данни.
5. Проектирана е учебна програма по дисциплината „Въведение в науката за данните“ и е изградена методология за обучение, адаптирана към нея.

Основният извод е, че за да бъдат изградени компетенции в областта на науката за данните, е важно да бъде организирано обучение, основано на компетенции.

ЗАКЛЮЧЕНИЕ

Днешният свят се характеризира с повсеместно навлизане на ИКТ във всички аспекти на живота. Огромни по обем, разнообразие и темп на растеж и промяна данни се събират, съхраняват и натрупват в компютърните системи. Вследствие на това организациите, които днес работят в условия на глобална икономика, конкуренция и динамични пазари, са изправени пред предизвикателството и на големите данни. Осъзнава се нуждата от изграждане на необходимата експертна инфраструктура за работа с данни с цел подпомагане на процесите на вземане на решение, трансформация и иновирание на важни процеси като производство, ценообразуване, доставки, маркетинг, анализ и прогноза на потребителското поведение и др. Полезното използване на натрупаните данни изисква значителна експертиза, както в областта на компютърните технологии, така и в области, свързани с анализ на данни, като математика и статистика, извличане на знания от данни, и способности за интерпретиране на резултатите от анализа, отнесени към предметната област на решавания проблем. Технологиите, създадени за обработката на големите данни позволяват на анализаторите, изследователите и бизнес потребителите да използват данни, които преди това са били недостъпни или неизползваеми, което води до подобряване на процесите на вземане на решения. Във връзка с гореизложеното се налага изводът, че специалистите по анализ на големи данни се превръщат във все по-ценен ресурс за разбиране на социалните и икономически процеси в тяхната цялост.

Настоящото монографично проучване представя актуално изследване в областта на големите данни и наука за данните. Основната цел на монографията е свързана със създаването на план за обучение по наука за данните като подход за преодоляване на дигиталното разделение в епохата на големите данни. За постигане на поставената цел са осъществени следните логически свързани с нея изследователски задачи:

- анализ и дискусия на същността, ролята и нарастващото значение на големите данни и специалистите, които ги анализират и извличат полезно знание от тях.
- анализ на предизвикателствата, пред които са изправени специалистите по данни в сферата на науката за данните;

- изграждане на професионален профил на специалиста по данни от гледна точка на необходимите знания, умения и компетенции, които той трябва да притежава, с цел обработка и анализ на големи данни;
- изследване и анализ на търсенето (от страна на бизнес организациите) и предлагането (от страна на обучаващите организации) на специалисти в областта на науката за данните в България;
- изследване и анализ на предварителната подготовка на бъдещите специалисти по данни (потенциалните кандидати за магистърска специалност в областта на науката за данни) от гледна точка на тяхното аналитично и логическо мислене;
- разработване на модел на учебен план, учебна програма и методология за обучение по наука за данните, които се базират върху формирането на компетенции за работа с данни.

Процесът на изграждане на учебен план за обучение по наука за данните е многостъпков и обхваща сериозни проучвания на гореизложените ключови фактори по отношение на компетенциите на специалистите по данни, както и тяхното търсене и предлагане на пазара. Магистърската програма по наука за данните предвижда овладяването на задълбочени теоретични познания и усвояването на разнообразни практически умения, свързани с приложения на методи и техники в широк кръг от направления на информатиката и информационните технологии за извличане на скрити смислени модели и закономерности в големи масиви от неструктурирани данни. Основен акцент представлява дисциплината „Въведение в науката за данните“ и изградената методология за обучение, адаптирана към нея. Тя има за цел да запознае студентите с фундаменталните концепции в съвременната теория на науката за данните и основни методи и технологии за извличане, организиране и анализ на данни, да развие на критичното и аналитично мислене, включително доказателства, оценка, анализ и синтез на множество източници. Тази дисциплина отговаря на съвременните нужди за обучение в сравнително новата област наука за данните, свързана с обработката и анализа на данните. Тя може да намери приложение като въведение в предметната област, също така за повишаване на грамотността на студентите при работа с информация, данни и големи данни, дори за нетехнически специалности.

ОСНОВНИ ПРИНОСИ В МОНОГРАФИЯТА

Приноси на теоретично равнище

1. Систематизирани са основни концепции по отношение на проблемите и предизвикателствата, стоящи пред бизнес организациите, специалистите по данни и обучаващите институции в ерата на големите данни.
2. Изграден е професионален профил на специалиста по данни, който описва необходимия набор от знания и умения, които трябва да притежава съвременния специалист по данни.
3. Разработен е оперативен компетентностен модел на професионални компетенции, включващ четири групи умения: „твърди“, „меки“, аналитични и етични.
4. Разработен е универсален въпросник, който изследва аналитичното мислене и чиито въпроси засягат аналитичните умения, прилагането на абстрактно и логическо мислене, различаването на факти от мнения, както и прилагането на количествено мислене.
5. Моделирана е система за формиране на професионални компетенции в областта на науката за данните.

Приноси на научно-приложно равнище

1. В резултат на проведеня литературен обзор и теоретичен анализ са изведени обобщения и са формулирани основните работни понятия на изследването.
2. Осъществен е анализ на входящите умения на потенциалните кандидати за магистърска програма по „Анализ и управление на данни“. Входящите умения на кандидатите служат при изграждането на учебен план, формиращ задължителните компетенции на специалиста по данни.
3. Разработени са ориентирани към изграждане на компетенции в областта на науката за данните учебен план, каталог от компетенции, учебна програма и учебна методология.

Приложение 1. АНКЕТНА КАРТА

Проучване на степента на прилагане на аналитичност и критичност при вземане на решения

Тестът за оценка на аналитичните способности е единственият валидиран за България инструмент, който измерва уменията на кандидатите за разбиране и интерпретиране на текстове с различна степен на сложност, извеждане на логически умозаклучения и критично различаване на мнения от факти в разнородна по съдържание информация.

Тестът не изисква специална предварителна подготовка. Необходима е единствено тиха среда, в която да можете да се концентрирате и да покажете най-доброто от себе си. Тестът ще оцени уменията Ви чрез следните типове задачи:

- анализ на информацията;
- логически заключения;
- критично мислене;
- математическа логика.

Участието Ви е напълно анонимно. Предварително Ви благодарим за отделеното време!

<p>1. Последната завършена от Вас степен на образование е:</p> <p><input type="checkbox"/> средно образование</p> <p><input type="checkbox"/> ОКС „Професионален бакалавър“</p> <p><input type="checkbox"/> ОКС „Бакалавър“</p> <p><input type="checkbox"/> ОКС „Магистър“</p> <p><input type="checkbox"/> ОНС „Доктор“</p>	<p>2. Вие се обучавате по/сте завършили специалност</p> <p>(в случай че сте активен студент, запишете специалността, която следвате в момента; в случай че не сте студент, посочете последната завършена от Вас специалност):</p> <p>.....</p>
<p>3. Направете самооценка на притежаваните от Вас аналитични и логически умения:</p> <p><input type="checkbox"/> висока степен на владееене и прилагане</p> <p><input type="checkbox"/> средна степен на владееене и прилагане</p> <p><input type="checkbox"/> ниска степен на владееене и прилагане</p> <p><input type="checkbox"/> Не мога да преценя.</p>	<p>4. Вие сте специалист или се обучавате в професионално направление:</p> <p><input type="checkbox"/> Информатика и компютърни науки</p> <p><input type="checkbox"/> Обществени комуникации и информационни науки</p> <p><input type="checkbox"/> История и археология</p> <p><input type="checkbox"/> Национална сигурност</p> <p><input type="checkbox"/> Друго:</p> <p>.....</p>

Индекс на издръжката на живота		
	Хранителни стоки и	
	безалкохолни напитки	алкохолни напитки и тютюнови изделия
Година 1	94,9	90
Година 2	96,9	94,7
Година 3	97,4	98,4
Година 4	99	99,4
Година 5	100	100
Година 6	100,6	100,8
Година 7	102	102,7
Година 8	103	104,7
Година 9	101,7	106
Година 10	101,2	107,5

Легенда: година 5 е база = 100 индекс-пункта

5. Какъв основен извод можете да направите от данните, представени в таблицата?

- Цените на алкохолните напитки и тютюневите изделия се покачват с 10% между години 1 и 5.
- Цените на алкохолните напитки и тютюневите изделия през дадения период се покачват непропорционално спрямо цените на хранителните стоки и безалкохолните напитки.
- През разглеждания период са увеличени акцизите върху алкохолните напитки и тютюневите изделия.

6. През април 1994 г. по интернет е проведено социологическо проучване сред 1975 случайно подбрани респонденти за нагласите за изборите. 85% от респондентите са отговорили, че ще гласуват за СДС. Изборите печели БСП с 43% срещу 24% за СДС. Каква е причината за грешката в социологическото проучване?

- Извадката е малка.
- Извадката не е представителна.
- Анализът на данните е пристрастен.
- Резултатите са манипулирани съзнателно.

7. Г-н Иван Иванов Иванов, ЕГН 9402293254, неженен, завършил специалност „Компютърни науки“ в УниБИТ в София, кандидатства за работа в софтуерна компания в Австрия. Получава работата за пробен период от шест месеца, но му е отказана работна виза. Защо?

- заради семейното му положение
- заради противоречиви лични данни
- заради противоречиви данни за образованието
- заради възрастта

8. Машина произвежда за минута 100 броя от даден продукт. Ако в една кутия се опаковат 24 броя от продукта, колко кутии може да напълни машината за един час?

- 125
- 250
- 500
- 2500

9. Камион се движи със скорост 80 km/h. Колко километра ще пропътува за 30 min?

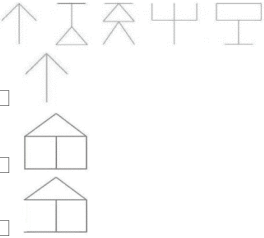
- 15 km
- 30 km
- 40 km
- 80 km
- никое от посочените разстояния

10. Някои мечки са джуджета. Всички джуджета обичат да рисуват. Следователно всички мечки обичат да рисуват.

- Правилно
- Неправилно

11. Всеки ромб е кръгъл. Всички ромбове са жълти. Следователно:

- жълтите ромбове имат кръгли ъгли
- съществуват ромбове с жълти ъгли
- съществуват кръгли жълти ъгли
- съществуват ромбове с кръгли ъгли
- ромбовете и ъглите са жълти и кръгли
- Няма верен отговор.

<p>12. Някои хора са марсианци. Марсианците са с три ръце. Следователно:</p> <p><input type="checkbox"/> някои хора имат по три ръце</p> <p><input type="checkbox"/> марсианците с две ръце понякога са хора</p> <p><input type="checkbox"/> марсианците са хора с три ръце</p> <p><input type="checkbox"/> хората с две ръце не са марсианци</p> <p><input type="checkbox"/> марсианците, които са хора, понякога са с три ръце</p>	<p>13. Всички котки са черни. Някои черни животни са ловки. Затова:</p> <p><input type="checkbox"/> всички черни животни са котки</p> <p><input type="checkbox"/> всички котки са ловки</p> <p><input type="checkbox"/> Няма верен отговор.</p>
<p>14. Изречението „Пресмятането на ум е трудно.“ изразява:</p> <p><input type="checkbox"/> факт</p> <p><input type="checkbox"/> мнение</p>	<p>15. Изречението „Някои хора се съмняват в това, че модерните комуникации ще направят живота ни по-лесен.“ изразява:</p> <p><input type="checkbox"/> факт</p> <p><input type="checkbox"/> мнение</p>
<p>16. Базирайки се на цифрите в английския език, с коя буква би следвало да продължи редицата OTTFFSS?</p> <p><input type="checkbox"/> E</p> <p><input type="checkbox"/> A</p> <p><input type="checkbox"/> W</p> <p><input type="checkbox"/> Нито една от посочените.</p>	<p>17. Каква трябва да бъде следващата фигура в редицата:</p>  <p><input type="checkbox"/></p> <p><input type="checkbox"/></p> <p><input type="checkbox"/></p>
<p>18. Ели, Крис и Станчо отседат в хотел. За стаята плащат общо 30 лв., като всеки плаща по 10 лв. Собственикът на хотела решава да им направи отстъпка от 5 лв., които дава на една от камериерките, казвайки <input type="checkbox"/> да им ги предаде. По пътя тя се замисля, че не може да раздели поравно 5 лв. на тримата, затова решава да прибере 2 лв. за себе си и да даде по 1 лв. на всеки от гостите. Така всеки е платил по $10 - 1 = 9$ лв. за стаята, 2 лв. са отишли при камериерката. Общо стават $3 * 9 + 2 = 29$ лв. Къде се губи последният лев?</p> <p><input type="checkbox"/> Камериерката го е прибрала.</p> <p><input type="checkbox"/> Задачата е грешно решена. <input type="checkbox"/> Върху сумата е наложен данък.</p>	
<p>19. В съд с формата на триъгълник има три мравки – по една във всеки от ъглите на триъгълника. Едновременно всяка от тях тръгва към някой от другите ъгли. Каква е вероятността, след като стигнат до ъгъла, към който са се отправили, отново във всеки ъгъл на триъгълника да има по точно една мравка?</p> <p><input type="checkbox"/> 100%</p> <p><input type="checkbox"/> 25%</p> <p><input type="checkbox"/> 50%</p> <p><input type="checkbox"/> 0%</p>	<p>20. Всички цветя в един букет, освен две, са рози; всички, освен две, са маргаритки и всички, освен две, са лалета. Колко цветя има в букета?</p> <p><input type="checkbox"/> 3</p> <p><input type="checkbox"/> 6</p> <p><input type="checkbox"/> 12</p>

21. Седем сестри пристигат в къщата си на село и всяка се захваща с някаква работа. Първата започва да чете любовен роман, втората – да пържи бухти, третата – да играе на карти, четвъртата – да решава кръстословица, петата – да пере, а шестата – да полива градината. Какво започва да прави седмата сестра?

- да гледа телевизия
- да яде бухти
- да играе на карти

22. Как бихте оценили своите аналитични и логически умения след решаването на задачите от теста?

1 2 3 4 5

Ниска степен на
владееене и прилагане



Висока степен на
владееене и прилагане

Благодарим Ви за участието!

Приложение 2. АЛГОРИТЪМ НА ОБРАБОТКА НА ДАННИТЕ

Авторките изказват своята благодарност на доц. д-р Магдалена Гарванова за проявения професионализъм и подкрепа при осъществяването на статистическата обработка на данните.

1. Рекодиране

На първия етап на обработка на данните всичките 17 айтема, участващи във формирането на четирите подскали (АУ, АЛМ, ФМ и КМ), бяха рекодирани, като код 0 беше приписан на грешните, а код 4 – на верните отговори, съобразно ключа към картата.

2. Формиране на бал

Беше формиран бал по всяка скала като сума от отговорите на въпросите, които влизат в нея.

3. Описателна статистика

В резултат на приложените техники за трансформация на данните се получи следната описателна статистика на отделните скали:

		Statistics			
		Аналитични умения	Абстрактно-логическо мислене	Факти от мнения	Количествено мислене
N	Valid	196	200	202	199
	Missing	8	4	2	5
Mean		10,90	10,30	4,50	15,46
Std. Deviation		4,70	3,77	3,01	4,97
Minimum		0	0	0	0
Maximum		20	16	8	24

4. Нормиране на скалите

С цел да се диференцират лицата с ниско, средно и високо ниво на аналитични умения, се направи нормиране на оценката на бала за четирите подскали поотделно. Средното равнище на всяка скала се получава, когато балът е в интервала на *средната стойност ± стандартното отклонение*. Съответно ниското равнище е под този интервал, а високото – над него. В резултат на приложеното нормиране бяха получени следните процентни разпределения:

Аналитични умения_gr					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ниска степен	29	14,2	14,8	14,8
	Средна степен	111	54,4	56,6	71,4
	Висока степен	56	27,5	28,6	100,0
	Total	196	96,1	100,0	
Missing	System	8	3,9		
Total		204	100,0		

Абстрактно-логическо мислене_gr					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ниска степен	26	12,7	13,0	13,0
	Средна степен	142	69,6	71,0	84,0
	Висока степен	32	15,7	16,0	100,0
	Total	200	98,0	100,0	
Missing	System	4	2,0		
Total		204	100,0		

Факти от мнения_gr					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ниска степен	46	22,5	22,8	22,8
	Средна степен	85	41,7	42,1	64,9
	Висока степен	71	34,8	35,1	100,0
	Total	202	99,0	100,0	
Missing	System	2	1,0		
Total		204	100,0		

Количествено мислене_gr					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ниска степен	24	11,8	12,1	12,1
	Средна степен	161	78,9	80,9	93,0
	Висока степен	14	6,9	7,0	100,0
	Total	199	97,5	100,0	
Missing	System	5	2,5		
Total		204	100,0		

5. Корелационен анализ (r на Пирсън)

Correlations					
		Анали- тични умения	Абстракт- но-логи- ческо мислене	Факти от мнения	Количест- вено мислене
Аналитични умения	Pearson Correlation	1	0,332*	0,133	0,368*
	Sig. (2-tailed)		0,000	0,065	0,000
	N		194	195	192
Абстрактно- логическо мислене	Pearson Correlation		1	0,061	0,296*
	Sig. (2-tailed)			0,388	0,000
	N			200	196
Факти от мнения	Pearson Correlation			1	0,100
	Sig. (2-tailed)				0,161
	N				198
Количестве- но мислене	Pearson Correlation				1
	Sig. (2-tailed)				
	N				

* Correlation is significant at the 0,01 level (2-tailed).

Корелационният анализ показва, че е налице положителна и умерена по сила статистически значима взаимовръзка между АУ, АЛМ и КМ, от една страна, както и между АЛМ и КМ, от друга страна. ФМ не корелира статистически значимо с нито една подскала от изследването. Допускането е, че ако във ФМ влизат повече айтеми, а не само два, някои тенденции във взаимовръзките между променливите ще имат статистически достоверен ефект (например между ФМ и АУ, където нивото на значимост на корелационния коефициент има гранични стойности – $r = 0,133$ при $p = 0,065$).

Приложение 3. КАТАЛОГ ОТ КОМПЕТЕНЦИИ КЪМ ПРОГРАМА „АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“

3.1. Общи компетенции

Наименование	Дефиниране на дейността
Планиране и организиране	Подготвя ясни краткосрочни и дългосрочни планове; подрежда приоритетите ефективно, наблюдава напредъка и осигурява изпълнението на следващите стъпки; прилага систематичен подход.
Практическо решаване на проблеми	Анализира и разбира проблема (ситуацията, казуса), като го разделя на по-малки, обозрими за обхващане елементи и части или проследява стъпка по стъпка произтеклите последици; организира и свързва методично елементи от проблема/ситуацията/казуса; проверява причините; систематично сравнява различни характеристики и аспекти; организира приоритетите на базата на аргументи, разглежда събитията, разположени хронологично във времето, както и причинно-следствените връзки или отношения „ако..., то...“.
Изготвяне на оптимизирани решения	Поставя фокуса върху спазването на закона, процедурите и правилата при подготовката на решения; ориентира решенията към максимален резултат с минимални разходи на ресурси.
Анализиране на данни и информация	Анализира данни и информация и прави изводи за очертаващи се тенденции; предлага и взема аргументирани решения, следващи логично от анализа.
Аналитично мислене	Подбира подходящите източници на информация; критично оценява и анализира точно и логично комплексни данни; анализира цялата събрана информация, когато взема решения и дава препоръки.
Умения за междумоментна устна комуникация/общуване	Осигурява информация така, че другите да могат да разберат ситуацията; умее да изслушва и разбира позициите на различните страни и да предприема необходимите действия своевременно и ефикасно; може да приеме убедено и доброволно преосмислянето и промяната на собствената си позиция с цел подобряване на крайния резултат; включва разбиране и предоставяне на информация на други хора на различни нива в организацията и извън нея (служители и клиенти) и желание да се учи от действията на другите и от своите собствени

Наименование	Дефиниране на дейността
	грешки.
Методичност при изпълнение на задачите	Организира ефективно задачите и времето си; изпълнява рутинната си работа съвестно и последователно; обръща внимание на подробностите и следи за спазването на процедурите.
Работа в екип и сътрудничество	Работи в сътрудничество с други колеги, а не самостоятелно или в конкуренция с тях; съобразява се със и прилага правилата на екипа; проявява способност да насърчава и подпомага останалите и да участва в открит, непрекъснат обмен на информация и мнения, различаващи се от останалите; екипът може да е в рамките на един отдел, дирекция или цялата организация.

3.2. Специфични компетенции

Компетенция	Описание	Задачи
Боравене с извадки от данни	Събиране и селектиране на набор от данни от популация чрез статистическа или друг вид процедура	Оценяване на данни, информация и цифрово съдържание
Докладване на резултати от анализ	Изготвяне на документи или презентации с цел представяне на резултатите от проведен изследователски и аналитичен проект, като се посочват процедурите и методите за анализ, както и потенциалните интерпретации на резултатите	Докладване на факти; проверяване на данни; извършване на анализ на данни
Извършване на аналитични математически изчисления	Прилагане на математически методи и изчислителни технологии за извършване на анализи и решаване на проблеми	Извършване на маркшайдерски изчисления, на разчети за електрооборудване, на математически изчисления при борбата с вредителите, на свързани с работата изчисления; прилагане на аналитично мислене
Извършване на почистване на данни	Откриване и коригиране на повредени записи; осигуряване на структурираност на данните съгласно изискванията	Прецизиране на данните
Нормализиране на данни	Редуциране на данните (нормални форми), за да се постигне минимизация на зависимостите; премахване на излишните данни за постигане на консистентност на данните	Установяване на процеси за данни

Компетенция	Описание	Задачи
Предоставяне на визуална презентация на данни	Създаване на визуални изображения на данни, като графики и диаграми, за по-лесно разбиране	Прилагане на техники за визуално представяне
Прилагане на процеси за качество на данните	Прилагане на техники за анализ на качеството, валидиране и верификация	Използване на бази данни; прилагане на стандарти за качество; записване на производствените данни за контрол на качеството
Проектиране на схема за база данни	Проектиране на бази данни с цел създаване на логически свързани групи обекти, като таблици, колони и процеси	Управление на база данни; създаване на инструменти за разработване на бази данни
Разработване на приложения за обработване на данни	Създаване на персонализиран софтуер за обработка на данни чрез подходящ език за програмиране	
Събиране на ИКТ данни	Прилагане на подходящи методи	Събиране на данни; управление на системи за събиране на данни
Интерпретиране на данни	Анализ на събраните данни от източници, като пазарни данни, научни доклади и изследвания, изисквания на клиентите и въпросници, които са актуални, за да се оценят развитието и иновациите в разглежданата област	Тълкуване на сеизмични данни; разчитане на данни, свързани с производството на храни; тълкуване на данни за извличане; използване на методи за анализ на логистични данни; тълкуване на данните за автоматичното разпределяне на повижаванията; оценяване на надеждността на данните; анализиране на данни за околната среда; тълкуване на геофизични данни
Управление на системи за събиране на данни	Разработване и управление на методи и стратегии, използвани за повишаване на качеството на данните и статистическата ефективност при събирането им, за да се гарантира, че събраните данни са оптимизирани за по-нататъшна обработка	Управление на данни; събиране на ИКТ данни; управление на колекции
Установяване на процеси за данни	Използване на ИКТ и инструменти за прилагане на математически алгоритмични или други процеси за обработка на данни с цел извличане на смислена информация	Нормализиране на данни

Приложение 4. УЧЕБЕН ПЛАН НА МАГИСТЪРСКА ПРОГРАМА „АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“

**4.1. За студенти, завършили ОКС „Бакалавър“
в същото професионално направление**

**УНИВЕРСИТЕТ ПО БИБЛИОТЕКОЗНАНИЕ
И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ
Факултет „Информационни науки“
У Ч Е Б Е Н П Л А Н**

Област на висше образование: **4. Приложни науки, математика и информатика**

Професионално направление: **4.6. „Информатика и компютърни науки“**

Специалност: **АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ
(за студенти, завършили ОКС „Бакалавър“
в същото професионално направление)**

Форма на обучение: **РЕДОВНА**

Образователно-квалификационна степен: **МАГИСТЪР**

Срок на обучение: **1 години (3 семестъра)**

Професионална квалификация:

АНАЛИЗАТОР И МЕНИДЖЪР НА ДАННИ И ИНФОРМАЦИЯ

Обучението по учебния план започва през: **м. октомври 2018 г.**

Код на учебния план:

УЧЕБЕН ПЛАН

№	Дисциплина	Се- мес- тър	Кре- дита ECTS	Общо	Аудиторна застот				Извънауди- торна застот	Форма на изпитване
					АО	Л	У	П		
1-ВИ СЕМЕСТЪР										
1.	Въведение в науката за данни	1	5	125	60	30	30	0	65	И
2.	Статистика	1	5	125	60	30	30	0	65	И
3.	Хранилища на данни (<i>Data warehousing</i>)	1	5	125	60	30	30	0	65	И
4.	Извличане на данни (<i>Data mining</i>)	1	5	125	60	30	30	0	65	И
5.	Визуализация на данни. Ергономия	1	5	125	60	30	30	0	65	И
6.	Избираема дисциплина	1	5	125	60	30	30	0	65	И
	Общо за 1-ви семестър		30	750	360	180	180	0	390	6И
2-РИ СЕМЕСТЪР										
1.	Анализ на големи данни	2	5	125	60	30	30	0	65	И
2.	Архитектура на приложение	2	5	125	60	30	30	0	65	И
3.	Въведение в облачните технологии	2	5	125	60	30	30	0	65	И

№	Дисциплина	Семес- тър	Кре- дита ECTS	Общо	Аудиторна заетост				Извънауди- торна заетост	Форма на изпитване
					АО	Л	У	П		
4.	Избираема дисциплина	2	5	125	60	30	30	0	65	И
5.	Избираема дисциплина	2	5	125	60	30	30	0	65	И
6.	Избираема или факултативна дисциплина	2	5	125	60	30	30	0	65	И/Т
Общо за 2-ри семестър			30	750	360	180	180	0	390	5И + И/Т
3-ТИ СЕМЕСТЪР										
1.	Научноизследователски проект	3	15	375	0	0	0	0	375	Зпр
2.	Магистърска теза	3	15	375	0	0	0	0	375	Змт
Общо за 3-ти семестър			30	750	0	0	0	0	750	Зпр + Змт

ИЗБИРАЕМИ ДИСЦИПЛИНИ

№	Дисциплина	Семес- тър	Кре- дита ECTS	Об- що	Аудиторна заетост				Извънауди- торна заетост	Форма на из- пита- ване
					АО	Л	У	П		
1-ВИ СЕМЕСТЪР										
1.	Поведенческа икономика	1	5	125	60	30	30	0	65	И
2.	Технологии за данни	1	5	125	60	30	30	0	65	И
3.	Разработка и управление на изискванията	1	5	125	60	30	30	0	65	И
2-РИ СЕМЕСТЪР										
1.	Разкриване на заплахи и измами	2	5	125	60	30	30	0	65	И
2.	ERP системи	2	5	125	60	30	30	0	65	И
3.	Анализ и управление на данни в ДА	2	5	125	60	30	30	0	65	И
4.	Моделиране и симулиране в графична среда	2	5	125	60	30	30	0	65	И
5.	Създаване и управление на проекти	2	5	125	60	30	30	0	65	И
6.	Интернет на нещата	2	5	125	60	30	30	0	65	И
7.	Управление на отношенията с клиенти (CRM)	2	5	125	60	30	30	0	65	И

ФАКУЛТАТИВНИ ДИСЦИПЛИНИ

Дисциплина	Семес- тър	Кре- дита ECTS	Об- що	Аудиторна заетост				Извънаудиторна заетост	Форма на изпитва- не
				АО	Л	У	П		
2-РИ СЕМЕСТЪР									
Количествени методи	2	5	125	60	30	30	0	65	Т

ОБОБЩЕНА ИНФОРМАЦИЯ ЗА УЧЕБНИЯ ПЛАН

Кредити и хорариум на дисциплините

Семестър	Кредити ECTS	Хорариум						Седмична аудиторна заетост (ч)
		Общо	Аудиторна заетост				Извънаудиторна заетост	
			Общо АО	Л	У	П		
1	30	750	360	180	180	0	390	24
2	30	750	360	180	180	0	390	24
3	30	750	0	0	0	0	750	0
Общо	90	2250	720	360	360	0	1530	

Брой дисциплини и брой форми на контрол

Семестър	Брой дисциплини			Брой изпити	Брой текущи/продължаващи оценки
	Общо	Задължителни	Избираеми / факултативни		
1	6	5	1	6	0
2	6	3	3	5	1
3	Зпр + ЗМТ	Зпр + ЗМТ	0	Зпр + ЗМТ	0
Общо	12 + Зпр + ЗМТ	8 + 1 Зпр + 1 ЗМТ	4	11 + 1 Зпр + 1 ЗМТ	0
Общо дисциплини	14	10	4	13	1

Легенда:

Аудиторна заетост в семестъра: **АО** – общ брой; от тях: **Л** – лекции, **У** – упражнения, **П** – учебна и специализираща практика.

Извънаудиторната заетост включва: консултации с преподавател; самостоятелни упражнения, подготовка в библиотеки, посещение на публична лекция и др.; самостоятелна работа (написване на публикация, реферат, курсова работа, друг вид разработка); самостоятелна подготовка.

Форма на изпитване – със стойности: **И** – изпит; **Т** – текуща оценка (формира се от: текущ контрол – тестове, колоквиум, домашна работа и т.н.); **Зпр** – Защита на проект; **ЗМТ** – Защита на магистърска теза.

1 ECTS = 25 учебни часа

Форми на завършване на обучението

Обучението завършва със защита на магистърска теза (**Зм**).

Забележки

* Списъкът на избираемите и факултативните дисциплини се актуализира в края на всяка учебна година с решение на ФС.

* Дневната аудиторна заетост е отношението между часовете обща аудиторна заетост (АО) и броя на дните, предвидени за очните занятия, в часове, по семестри. Тя не може да надвишава 10 часа.

4.2. За студенти, завършили ОКС „Бакалавър“ в друго професионално направление

**УНИВЕРСИТЕТ ПО БИБЛИОТЕКОЗНАНИЕ
И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ
Факултет „Информационни науки“
У Ч Е Б Е Н П Л А Н**

Област на висше образование: **4. Приложни науки, математика и информатика**

Професионално направление: **4.6. „Информатика и компютърни науки“**

Специалност: **АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ
(за студенти, завършили ОКС „Бакалавър“ в друго професионално направление)**

Форма на обучение: **РЕДОВНА**

Образователно-квалификационна степен: **МАГИСТЪР**

Срок на обучение: **1 години (3 семестъра)**

Професионална квалификация: **АНАЛИЗАТОР И МЕНИДЖЪР
НА ДАННИ И ИНФОРМАЦИЯ**

Обучението по учебния план започва през: **м. октомври 2018 г.**

Код на учебния план:

УЧЕБЕН ПЛАН

№	Дисциплина	Семестър	Кредити ECTS	Общо	Аудиторна заетост				Извънаудиторна заетост	Форма на изпитване
					АО	Л	У	П		
1-ВИ СЕМЕСТЪР										
1.	Въведение в науката за данни	1	5	125	60	30	30	0	65	И
2.	Статистика	1	5	125	60	30	30	0	65	И
3.	Хранилища на данни (<i>Data warehousing</i>)	1	5	125	60	30	30	0	65	И
4.	Извличане на данни (<i>Data mining</i>)	1	5	125	60	30	30	0	65	И
5.	Визуализация на данни. Ергономия	1	5	125	60	30	30	0	65	И

6.	Математически анализ*	1	3	75	60	30	30	0	15	И
7.	Теория на вероятностите и статистика*	1	3	75	60	30	30	0	15	И
8.	Избираема дисциплина	1	5	125	60	30	30	0	65	И
Общо за 1-ви семестър			36	900	480	240	240	0	420	8И
2-РИ СЕМЕСТЪР										
1.	Анализ на големи данни	2	5	125	60	30	30	0	65	И
2.	Архитектура на приложения	2	5	125	60	30	30	0	65	И
3.	Въведение в облачните технологии	2	5	125	60	30	30	0	65	И
4.	Бази данни*	2	3	125	60	30	30	0	65	И
5.	Обектно ориентирано програмиране*	2	3	75	60	30	30	0	15	И
6.	Избираема дисциплина	2	5	75	60	30	30	0	15	И
7.	Избираема дисциплина	2	5	125	60	30	30	0	65	И
8.	Избираема или факултативна дисциплина	2	5	125	60	30	30	0	65	И/Т
Общо за 2-ри семестър			36	900	480	240	240	0	420	7И + И/Т
3-ТИ СЕМЕСТЪР										
1.	Научноизследователски проект	III	15	375	0	0	0	0	375	Зпр
2.	Магистърска теза	III	15	375	0	0	0	0	375	Змг
Общо за 3-ти семестър			30	750	0	0	0	0	750	Зпр + Змг

ИЗБИРАЕМИ ДИСЦИПЛИНИ

№	Дисциплина	Семес- тър	Кре- дита ECTS	Об- що	Аудиторна заетост				Извъ- науди- торна заетост	Форма на из- пит- ване
					АО	Л	У	П		
1-ВИ СЕМЕСТЪР										
1.	Поведенческа икономика	1	5	125	60	30	30	0	65	И
2.	Технологии за данни	1	5	125	60	30	30	0	65	И
3.	Разработка и управление на изискванията	1	5	125	60	30	30	0	65	И
2-РИ СЕМЕСТЪР										
1.	Разкриване на заплахи и измами	2	5	125	60	30	30	0	65	И
2.	ERP системи	2	5	125	60	30	30	0	65	И
3.	Анализ и управление на данни в ДА	2	5	125	60	30	30	0	65	И
4.	Моделiranje и симулиране в графична среда	2	5	125	60	30	30	0	65	И
5.	Създаване и управление на проекти	2	5	125	60	30	30	0	65	И
6.	Интернет на нещата	2	5	125	60	30	30	0	65	И
7.	Управление на отношенията с клиенти (CRM)	2	5	125	60	30	30	0	65	И

ФАКУЛТАТИВНИ ДИСЦИПЛИНИ

№	Дисциплина	Семес- тър	Кре- дита ECTS	Об- що	Аудиторна заетост				Извънаудиторна заетост	Форма на из- питване
					АО	Л	У	П		
2-РИ СЕМЕСТЪР										
	Количествени методи	2	5	125	60	30	30	0	65	Т

ОБОБЩЕНА ИНФОРМАЦИЯ ЗА УЧЕБНИЯ ПЛАН

Кредити и хорариум на дисциплините

Семес- тър	Кредити ECTS	Хорариум						Сем. ауди- торна зае- тост (ч)
		Общо	Аудиторна заетост				Извънауди- торна заетост	
			Общо АО	Л	У	П		
1	36	900	480	240	240	0	420	32
2	36	900	480	240	240	0	420	32
3	30	750	0	0	0	0	750	0
Общо	102	2550	960	480	480	0	1590	

Брой дисциплини и брой форми на контрол

Семестър	Брой дисциплини			Брой из- пити	Брой теку- щи/продължаващи оценки
	Общо	Задължи- телни	Избирае- ми/факултативни		
1	8	7	1	8	0
2	8	5	3	7	1
3	Зпр + ЗМТ	Зпр + ЗМТ	0	Зпр + ЗМТ	0
Общо	16 + Зпр + ЗМТ	12 + 1 Зпр + 1 ЗМТ	4	15 + 1 Зпр + 1 ЗМТ	0
Общо дис- циплини	18	14	4	17	1

Легенда:

Аудиторна заетост в семестъра: **АО** – общ брой; от тях: **Л** – лекции, **У** – упражнения, **П** – учебна и специализираща практика.

Извънаудиторната заетост включва: консултации с преподавател; самостоятелни упражнения, подготовка в библиотеки, посещение на публична лекция и др.; самостоятелна работа (написване на публикация, реферат, курсова работа, друг вид разработка); самостоятелна подготовка

Форма на изпитване – със стойности: **И** – изпит; **Т** – текуща оценка (формира се от: текущ контрол – тестове, колоквиум, домашна работа и т.н.); **Зпр** – Защита на проект; **ЗМТ** – Защита на магистърска теза.

1 ECTS = 25 учебни часа

Форми на завършване на обучението

Обучението завършва със защита на магистърска теза (**Зм**).

Забележки

* Списъкът на избираемите и факултативните дисциплини се актуализира в края на всяка учебна година с решение на ФС.

* Дневната аудиторна заетост е отношението между часовете обща аудиторна заетост (АО) и броя на дните, предвидени за очните занятия, в часове, по семестри. Тя не може да надвишава 10 часа.

Приложение 5. КВАЛИФИКАЦИОННА ХАРАКТЕРИСТИКА НА СПЕЦИАЛНОСТ „АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“

УНИВЕРСИТЕТ ПО БИБЛИОТЕКОЗНАНИЕ
И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ
1784 София, бул. „Цариградско шосе“ № 119
Имейл: unibit@unibit.bg / kbd_college@yahoo.com
Факултет „Информационни науки“

КВАЛИФИКАЦИОННА ХАРАКТЕРИСТИКА

Специалност: „Анализ и управление на данни“
Професионално направление: 4.6. ИНФОРМАТИКА И
КОМПЮТЪРНИ НАУКИ
Образователно-квалификационна степен: МАГИСТЪР
Катедра: Компютърни науки
Форма на обучение: редовна/задочна
Срок на обучение: 1 година

Анотация

Магистърската програма е предназначена за студенти, завършили ОКС „Бакалавър“ в УниБИТ и други висши училища, както и работещи в организации на позиции, изискващи работа с данни.

Магистърската програма осигурява знания, компетентности и умения за:

- организиране на големи информационни масиви в условията на разнообразни ИКТ платформи;
- организиране на събирането и осигуряването на качеството на данни;
- извличане на полезни знания от данните при използване на статистически и други аналитични методи;
- оценка на чувствителността и доверието към резултатите от анализ на данни;
- проектиране, разработване, изграждане и използване на компютърни приложения, предназначени за анализ и управление на данни в различни предметни области.

Знания

Студентите получават специализирани теоретични познания по всички аспекти, свързани с: изграждането на бизнес информационни

системи, ориентирани към управление, което е базирано на данни; софтуерни приложения, създадени за анализ на данни; статистически и интелигентни методи, модели и техники за извличане на знания от данни, а също и техники за визуализация на резултатите, ориентирани към лесно и интуитивно възприемане на резултатите от анализа на данни.

Компетенции

Завършилите студенти притежават следните компетенции:

- използват съвременни методи за анализ и управление на данни;
- разработват и реализират успешни стратегии в областта на организация на процесите по получаване, верификация, структуриране, съхраняване и управление на информационна ресурси;
- анализират данни и правят изводи въз основа на данни с помощта на компютърни технологии;
- използват различни платформи за развиване на нови софтуерни продукти и услуги в областта на анализа на данни;
- събират, анализират и интерпретират информация, необходима за изпълнителска, организационно-управленска, консултантска и научноизследователска дейност.
- използват нормативната база за изграждане на подпомагащи системи и институции;
- формират информационно поведение и информационна култура в обществото;
- анализират и използват специални информационни масиви и др.

Умения

Студентите придобиват ключови умения за:

- интегрално прилагане в работата си на придобитите знания за анализ и управление на данни чрез информационни технологии;
- ефективна работа по анализ и планиране на развитието на информационните ресурси;
- решаване на проблеми в областта на управлението на данни;
- работа в екип, формулиране и вземане на самостоятелни решения;
- обучение през целия си живот и ефективно комуникиране на знания, технологии и добри практики в обществото и др.

Перспективи за реализация

Успешно завършилите магистри имат задълбочени аналитични знания и умения и могат успешно да се реализират във всякакви институции, нуждаещи се от знания, компетентности и умения за организиране, управление и анализ на данни. Институции като банки, застрахователни дружества, телекоми и др. са само малка извадка от потенциалните работодатели на тези специалисти.

СПИСЪК НА РЕЛЕВАНТНИТЕ ДЪЛЖНОСТИ В НАЦИОНАЛНАТА КЛАСИФИКАЦИЯ НА ПРОФЕСИИТЕ И ДЪЛЖНОСТИТЕ

Код		Наименование
1330	6012	Мениджър, обработка на данни
1330	6005	Ръководител, информационно обслужване
2521	6001	Аналитик, бази данни
2529	6003	Експерт, извличане на данни
2529	6006	Специалист, обработка на данни
2529	6002	Експерт, информационно осигуряване

Условия за прием

Приемът на студенти се осъществява въз основа на успеха от дипломата им за завършена бакалавърска степен съгласно изискванията на Закона за висшето образование.

**Приложение 6. УЧЕБНА ПРОГРАМА
ПО ДИСЦИПЛИНАТА „ВЪВЕДЕНИЕ
В НАУКАТА ЗА ДАННИ (DATA SCIENCE)“**

**УНИВЕРСИТЕТ ПО БИБЛИОТЕКОЗНАНИЕ
И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ**
1784 София, бул. „Цариградско шосе“ № 119
Имейл: unibit@unibit.bg

Утвърждавам
Ректор:
(проф. д-р И. Петева)

У Ч Е Б Н А П Р О Г Р А М А

Факултет

Факултет „Информационни науки“

Катедра

Компютърни науки

Професионално направление

4.6. Информатика и компютърни науки

Специалност

Анализ и управление на данни

ОПИСАНИЕ

1. Учебни резултати за курса – реферат на дисциплината, усвоени знания, умения, компетенции (цели)

Цели на дисциплината

- Получаване на теоретични знания – цялостна представа за състоянието, обхвата и ролята на *Data Science* в съвременното общество.
- Изграждане на практически умения – раглеждане на казуси, позволяващи определяне на задачите за решаване на проблем, изискващ компетентност от областта на *Data Science*.

Връзка с други учебни дисциплини

Дисциплината изгражда рамката на магистърската програма и очертава мястото на всяка отделна дисциплина като компонент *от Data Science*.

Тематичен обхват

Дисциплината обхваща широк кръг от теми, от история на еволюцията на разбирането за извличане на знания от данни до ролята на конкретни научни области. Разгледани са също ползите за обществото и социалните измерения на *Data Science*.

Очаквани резултати

В резултат от изучаването на дисциплината се очаква да се изгради цялостен поглед върху предизвикателствата, технологиите, научните резултати и ползите за обществото от съвременните достижения на *Data Science*.

2. Начин на преподаване

Лекции: Аудиторно – основни концепции

Упражнения: Разглеждане и решаване на казуси

3. Предварителни изисквания (знания и умения от предходно обучение) и изисквания за други (едновременни) курсове

Математика, Статистика, Базисни данни, Програмиране

4. Съдържание на курса

№	Тема	Лекции	Упражнения
1	Въведение в дисциплината	3	
2	Еволюция на технологиите за извличане на полезни знания от данни	3	3
3	Еволюция на компютърните технологии, свързани със събиране, съхраняване, достъп до и извличане на знания от данни	3	3
4	Статистика – наука за анализ на данни	3	3
5	Големи данни – предизвикателства и ползи	3	3
6	Хранилища на данни (<i>Data warehousing</i>) – задачи и приложимост	3	4
7	Анализ на данни (<i>Data Mining</i>) – задачи и приложимост	3	4
8	Машинно обучение	3	4
9	Езици и технологии (<i>R, Python, Minors</i>)	3	3
10	Социални аспекти на <i>Data Science</i> . Филантропия на данни	3	3

5. Планирани учебни дейности и методи на преподаване (общо описание)

Интерактивни лекции, дискусии

6. Методи и критерии на оценяване

Проект, тест

7. Език на преподаване

Български

Приложение 7. УЧЕБНА МЕТОДОЛОГИЯ

МАГИСТЪРСКА ПРОГРАМА: „АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE) ДИСЦИПЛИНА: ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ ЛЕКЦИЯ 1

Въведение в дисциплината

Цел: Запознаване на студентите с целта, обхвата и съдържанието на магистърската програма и дисциплината „Въведение в науката за данните“.

Съдържание:

1. Въведение

• *Тема 1.* Обем на данните (Volume). Взаимовръзка между обем на данни и информационни технологии. Исторически бележки

○ *Пример 1:* *Холерит: преброяване на населението в САЩ през 1893 г., изобретяване на перфокартите и сортировъчните машини.*

○ *Пример 2:* *Емануел Голдберг, Карл Цайс – Йена, Rapid Selector*

• *Тема 2.* Динамика (Velocity). Развитие на комуникационните технологии: от 1800 г. до 21. век.

○ *Пример 3:* *Комуникация, вземане на управленски решения в Източноиндийската компания*

• *Тема 3.* Разнообразие (Variety). Мултимедия, естествени езици
Обобщение: „3V“ модел на Gartner за големи данни

2. Съдържание на магистърската програма по Data Science

Цел на програмата

- Тематичен обхват
- Очаквани компетентности

3. Съдържание на дисциплината

- Теми
- Теория
- Практика
- Изпитване
- Оценяване

**МАГИСТЪРСКА ПРОГРАМА:
„АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE)
ДИСЦИПЛИНА:
ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ
ЛЕКЦИЯ 2**

Еволюция на технологиите за извличане на знания от данни

Цел: Запознаване на студентите с историята от установяването на принципите на научно познание и появата на научни дисциплини, пряко свързани с извличане на знания от данни.

Съдържание:

1. Преглед на основите на емпиричната и абстрактната наука – „източен“ и „западен“ път. Науката през вековете
2. Обособяване на съвременните научни направления. Природни науки, математика, статистика, социални науки
3. Процес на преобразуване на данни в знание:
 - Данни от наблюдения
 - Данни от контролирани експерименти
 - Построяване на модел
 - Верификация на модела
 - Използване на модела за прогнозиране
4. Машинно обучение (Machine Learning)
 - Статистическо машинно обучение
 - Deep Learning
 - Ръководено (Supervised) обучение
 - Неръководено (Unsupervised) обучение
5. Допълнителна тема: Управление на знанията. Теории за преобразуване на субективно (tacit) знание в записано (explicit)

**МАГИСТЪРСКА ПРОГРАМА:
„АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE)
ДИСЦИПЛИНА:
ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ
ЛЕКЦИЯ 3**

Еволюция на компютърните технологии, свързани със събиране, осигуряване на достъп и извличане на данни

Цел: Преглед на технологиите, свързани с данни, от 1950 г. насам и очертаване на тенденциите в развитието им.

Съдържание:

1. Етапи в развитието на компютърните технологии
2. Ранен етап – от машини, подпомагащи изчисленията, към машини за обработка на данни
3. Данни, качество на данните, представяне на данните в компютърните системи
4. Бази данни – основни идеи и изисквания. Йерархичен и мрежов модел
5. Бази данни – релационен модел и езикът SQL
6. Интернет
7. Ролята на проблема „bug 2000“ за трансформация на базите данни
8. 21 век: хетерогентни данни, облачни приложения, разпределени бази данни, социални мрежи

**МАГИСТЪРСКА ПРОГРАМА:
„АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE)
ДИСЦИПЛИНА: ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ
ЛЕКЦИЯ 4**

Статистика – наука за анализ на данни

Цел: Преглед на основните задачи, които се решават с помощта на методите на статистиката.

Съдържание:

1. Науката статистика – кратък исторически преглед
2. Основни задачи: построяване на модел, проверка на хипотези
3. Анализ на данни от контролирани експерименти
4. Анализ на данни от наблюдения
5. Параметрични и непараметрични методи
6. Основни направления:
 - „Честотна“ статистика
 - Бейсова статистика
7. Основни техники
 - Регресия
 - Класификация
 - Прогнозиране

**МАГИСТЪРСКА ПРОГРАМА:
„АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE)
ДИСЦИПЛИНА: ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ
ЛЕКЦИЯ 5**

Големи данни – предизвикателства и ползи

Цел: Запознаване с предизвикателствата, породени от масовото навлизане на компютърните технологии в живота на хората, което е свързано с регистрирането на всички факти в компютърни бази данни, известни като Big Data Challenge.

Съдържание:

1. Дефиниции на „големи данни“: „3V“, „5V“ и много V-та
2. Технологии за обработка на големи данни
3. Наука за данните
 - Трансдисциплинарен характер
 - Компетентности
4. Големи данни и иновации
5. Глобализация
6. Предизвикателства
7. Възможности
8. Етапи в еволюцията на организациите по отношение на реализирането на възможностите от използване на големи данни

**МАГИСТЪРСКА ПРОГРАМА:
„АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE)
ДИСЦИПЛИНА: ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ
ЛЕКЦИЯ 6**

Хранилища на данни (Data Warehouses) – задачи и приложимост

Цел: Запознаване с технологиите, структурата от данни и организацията на приложения, ориентирани към подпомагане на процеса на вземане на решения, известни като „бизнес интелигентност“, чието ядро са хранилищата на данни.

Съдържание:

1. Холистичен подход към решаването на бизнес проблеми.
Многомерни модели
2. Кубове с данни
3. Йерархии на димензиите
4. Процес ETL (extract-transform-load)
5. Изграждане на хранилища на данни:
 - Подход на Инмол
 - Подход на Кембъл
6. Визуализация
7. Технологии: MS Power BI, Tableau, Click

**МАГИСТЪРСКА ПРОГРАМА:
„АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE)
ДИСЦИПЛИНА: ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ
ЛЕКЦИЯ 7**

Анализ на данни (Data Mining) – задачи и приложимост

Цел: Запознаване с основните категории задачи, които се решават с инструментариума на Data Mining.

Съдържание:

1. Предварителна обработка:
 - Почистване на мръсни данни
 - Прогнозиране на липсващи данни
 - Дискретизация
 - Генерализация
 - Редуциране на размерността и обема
2. Характеризация и сравняване
3. Класификация и прогнозиране

**МАГИСТЪРСКА ПРОГРАМА:
„АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE)
ДИСЦИПЛИНА: ВЪВЕДЕНИЕ В НАУКАТА
ЗА ДАННИТЕ
ЛЕКЦИЯ 8
Машинно обучение**

Цел: Запознаване с основните категории задачи, които се решават с инструментариума, влизащ в категорията Machine Learning.

Съдържание:

1. Статистическо машинно обучение (Statistical Machine Learning):
построяване на модели, описващи изследваните обекти
2. Deep Learning: обучаване на „робот“ (компютърно приложение)
да взема решения. Изкуствени невронни мрежи
3. Контролирано машинно обучение (Supervised Machine Learning)
4. Неконтролирано машинно обучение (Unsupervised Machine Learning)
5. Процес на обучение – training and testing samples

**МАГИСТЪРСКА ПРОГРАМА:
„АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE)
ДИСЦИПЛИНА: ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ
ЛЕКЦИЯ 9**

Езици и технологии (R, Python, Data Miners)

Цел: Запознаване със съвременните езици, използвани за анализ на данни, със средите, изградени около тях, и с технологиите за подпомагане на дейностите по анализ на данни.

Съдържание:

1. Езикът R и обществото, поддържащо библиотеките с ресурси – програми за анализ на данни, RMiner
2. Езикът Python, библиотеките Panda и Anaconda
3. Rapid Miner, Knime, Orange и др.

**МАГИСТЪРСКА ПРОГРАМА:
„АНАЛИЗ И УПРАВЛЕНИЕ НА ДАННИ“ (DATA SCIENCE)
ДИСЦИПЛИНА: ВЪВЕДЕНИЕ В НАУКАТА ЗА ДАННИТЕ
ЛЕКЦИЯ 10**

Социални аспекти на Data Science. Филантропия на данни

Цел: Запознаване на студентите с понятието „филантропия“ в контекста на науката за данни и поставяне на акцент върху значението на публично-частното партньорство при обмена на данни.

Съдържание:

1. Data Science и организационната ефективност
2. Значение на публично-частното партньорство
3. Филантропията на данни като отговор на социалната отговорност на частните корпорации, притежаващи данни, и допълнителни ресурси към обществото

СПИСЪК НА ИЗПОЛЗВАНИТЕ СЪКРАЩЕНИЯ

АМ	Абстрактно мислене
АУ	Аналитични умения
БИС	Бизнес интелigentни системи
ВAAУ	Въпросник за анализ на аналитичните умения
ВСУ	Варненски свободен университет
ИКТ	Информационни и комуникационни технологии
ИСОК	Информационната система за оценка на компетенциите
ИТ	Информационни технологии
КМ	Количествено мислене
МОН	Министерство на образованието и науката
НАОА	Национална агенция за оценяване и акредитация
НБУ	Нов български университет
ПН	Професионално направление
СУ	Софийски университет
СУКО	Система за управление на качеството на обучение
ТУ	Технически университет
Уни-	Университет по библиотекознание и информационни
БИТ	технологии
УНСС	Университет за национално и световно стопанство
ФМ	Факти от мнения
ADV	Advanced Data Visualization
AI	Artificial Intelligence
ССТST	California Critical Thinking Skills Test
CDO	Chief Data Officer
DB	Data Base
DL	Deep Learning
DQP	Degree Qualification Profile
EDA	Exploratory Data Analysis
EDM	Educational Data Mining
EDSA	European Data Science Academy
ESCO	European Skills Competences and Occupations (Европейска комисия за оценка на уменията, компетентностите, квалификациите и професиите)

GDPR	General Data Protection Regulation
JSON	JavaScript Object Notation
IDC	International Data Corporation
LINQ	Language-Integrated Query
LDAP	Lightweight Directory Access Protocol
MDX	Multidimensional Expressions
MOOC	Massive Open Online Courses
NoSQL	Not only SQL
OLAP	Online Analytical Processing
ROC	Receiver Operating Characteristic
SQL	Structured Query Language
WGCTA	Watson-Glaser Critical Thinking Appraisal
XML	eXtensible Markup Language

ТЕРМИНОЛОГИЧЕН РЕЧНИК

Андрагогика
(андрагогика)

Раздел от теорията на обучението, разкриваща специфични закономерности при усвояването на знания и умения от възрастни хора на учебния материал, а също така и особености за ръководството на тази дейност от страна на професионален педагог.

Големи данни
(*Big Data*)

Данни с много голям обхват, структурирани или неструктурирани, събрани от значителен брой източници, които се обработват и анализират допълнително от машини, за да се извлекат и разберат по-добре заложените в тях тенденции и модели.

Данни
(*Data*)

Неструктурирани факти за обекти, които се съхраняват в „суров вид“, без да се използват.

Единица за обучение
(*Unit of Learning*)

Абстрактно понятие, използвано за обозначаване на курс, модул, урок и т.н.

Наука за данните
(*Data Science*)

Комбинация от техники, знания и умения, прилагани върху данните, за да се идентифицират скрити знания, които може да се използват за насочване на екипи при вземане на важни решения.

Специалист по данни
(*Data Scientist*)

Практик, който събира необработени данни и използва своите умения, знания и опит в областта на анализа, за да ги преобразува в полезна информация.

Дескриптивен анализ
(*Descriptive Analytics*)

Анализ, свързан основно с установяването на някакви присъщи характеристики и връзки, определящи състоянието на наличните до момента данни и поведението на процесите, които те описват.

Диагностичен анализ
(*Diagnostic Analytics, Inquisitive Analytics*)

Проучвателен анализ на съществуващите или допълнителни данни, който има за цел изследване и оценяване на влиянието на различни фактори и причини върху дадено следствие.

Дигитално неравенство
(Digital Divide)

Социално-икономическата разслоеност на обществото във връзка с достъпа до компютърните технологии и интернет. Това е вид неравенство, при което определена социална група има привилегиата да ползва информационно-комуникационни технологии, които са недостъпни за друга част от населението. Тази разслоеност може да се дължи на физическа невъзможност за достъп до компютърните технологии и/или на липса на знания и умения за боравене с тях.

Дигитално неравенство по повод на големи данни
(Big Data Divide)

Това е разделение между онези, които са в състояние да извлекат и използват непредвидими и необясними констатации от големите данни, и лицата, чието съществуване е засегнато от произтичащите от това решения.

Превантивен анализ
(Pre-emptive Analytics)

Анализ, свързан с възможността за предприемане на предпазни мерки за събития, които могат да имат нежелани последици върху организациите, например идентифициране на възможните опасности и препоръчване на смекчаващи стратегии далеч във времето.

Предписващ анализ
(Prescriptive Analytics)

Изследване на различните алтернативи във връзка с прогнозираните резултати, за да се подпомогнат анализаторите да оптимизират своя подход за постигане на най-добър отговор или действия, необходими за постигане на бизнес целите.

Прогностичен анализ
(Predictive Analytics)

Анализ, отнасящ се до обосноваване и конструиране на различни форми на научни предвиждания, разкриване на перспективи, но също и на възможни опасности и проблеми в бъдещото изменение на данните и параметрите на процесите въз основа на наблюдаваните им стойности и динамика до момента.

СПИСЪК НА ФИГУРИТЕ

Фиг. 1.1. „5V“ модел на големите данни	24
Фиг. 1.2. Жизнен цикъл на данните	29
Фиг. 1.3. Наука за данните – обобщаваща рамка на дефинициите	41
Фиг. 1.4. Облак от етикети	43
Фиг. 1.5. Данни – информация – знание	49
Фиг. 1.6. Дигитално разделение в сферата на големите данни и науката за данните	65
Фиг. 1.7. Цикъл на развитие на науката за данните	67
Фиг. 2.1. Триада „знания – умения и навици – компетенции“	75
Фиг. 2.2. Модел на компетентност „айсберг“	77
Фиг. 2.3. Приложение на видовете анализ.....	80
Фиг. 2.4. Умения на специалиста по данни	82
Фиг. 2.5. Профил на основните умения на специалиста по данни.....	89
Фиг. 2.6. Нетехнически аналитични умения	91
Фиг. 2.7. Задължителни умения в професионалния профил на специалиста по данни според работодателите.....	111
Фиг. 2.8. Търсени от бизнеса технически умения на специалиста по данни	112
Фиг. 3.1. Йерархия на мисловните умения	250
Фиг. 3.2. Групи компетентности в областта на науката за данните в образователни програми с общообразователен и изследователски профил	143
Фиг. 3.3. Групи компетентности в областта на науката за данните в образователни програми с бизнес профил	144
Фиг. 3.4. Процесът на разработване на въпросника	151

Фиг. 3.5. Процентно разпределение на респондентите от целевата и контролната група	156
Фиг. 4.1. Обща структура на триъгълника на дейностите.....	165
Фиг. 4.2. Систематичен подход за разработване и редизайн на магистърска програма, базирана на компетенции	168
Фиг. 4.3. Общ модел на система за формиране на професионални компетенции.....	170
Фиг. 4.4. Карта за надграждане на компетенциите от науката за информирането към науката за данните	177
Фиг. 4.5. Надграждане на уменията на инфоброкера	179
Фиг. 4.6. Процес на проектиране на учебна програма.....	181
Фиг. 4.7. Проектиране на учебен план – основни документи и ресурси	185
Фиг. 4.8. Мястото на учебната програма и учебната методология в обучение, основано на компетенции	197

СПИСЪК НА ТАБЛИЦИТЕ

Таблица 1.1. Дефиниции на „големи данни“.....	26
Таблица 1.2. Модели и технологии за големи данни.....	32
Таблица 1.3. Методи за анализ на големи данни.....	35
Таблица 1.4. Дефиниции за науката за данните.....	40
Таблица 1.5. Диаграми на Вен за обхвата на науката за данните.....	44
Таблица 1.6. Еволюция на анализа на данни.....	50
Таблица 1.7. Развитие на технологиите за работа с данни.....	51
Таблица 1.8. Дигитално неравенство в контекста на големите данни.....	60
Таблица 2.1. Сравнение на основните понятия.....	76
Таблица 2.2. Диференциация между професиите „специалист по данни“ и „анализатор на данни“.....	81
Таблица 2.3. Обхват на уменията на специалиста по данни.....	83
Таблица 2.4. Умения, прилагани от специалиста по данни – литературен преглед.....	84
Таблица 2.5. Компетентност и компетенции в науката за данните.....	87
Таблица 2.6. Основни задачи и необходими умения в науката за данните.....	89
Таблица 2.7. Етични умения при работа с лични данни.....	107
Таблица 2.8. Нива на експертност в областта на науката за данните.....	114
Таблица 2.9. Знания и опит, изисквани от работодателите.....	115
Таблица 3.1. Таксономия на Bloom.....	122
Таблица 3.2. Общи компетенции [16].....	129

Таблица 3.3. Компетенции на специалиста по данни според ESCO.....	132
Таблица 3.4. Специфични компетенции на специалиста по данни.....	133
Таблица 3.5. Основни курсове в програмите по наука за данните....	138
Таблица 3.6. Национални магистърски програми в областта на науката за данните.....	146
Таблица 3.7. Дисциплини, изграждащи учебните планове на магистърските програми в областта на науката за данните, предложени на национално равнище	147
Таблица 3.8. Дисциплини, изграждащи „твърди“ умения.....	148
Таблица 3.9. Разпределение на извадките	155
Таблица 3.10. Разпределение на респондентите по професионално направление	156
Таблица 3.11. Самооценка на аналитичните умения на техническата група	157
Таблица 3.12. Описателна статистика на отделните скали.....	158
Таблица 3.13. Прилагане на аналитични умения	158
Таблица 3.14. Прилагане на абстрактно и логическо мислене.....	158
Таблица 3.15. Различаване на факти от мнения	159
Таблица 3.16. Прилагане на количествено мислене.....	159
Таблица 3.17. Прилагане на АУ, АЛМ, ФМ и КМ от целевата група	159
Таблица 3.18. Корелационен анализ.....	160
Таблица 4.1. Матрица на пропуските [169]	172
Таблица 4.2. Дисциплини, формиращи учебния план на магистърска програма „Анализ и управление на данни“	176
Таблица 4.3. Дисциплини, надграждащи професионалния профил на информационния брокер	178
Таблица 4.4. Извадки от ЗВО, свързани с обучението в магистърска степен.....	182

Таблица 4.5. Разпределение на дисциплините по вид.....	188
Таблица 4.6. Дисциплини, включени в първия семестър на обучение	189
Таблица 4.7. Дисциплини, включени във втория семестър на обучение	189
Таблица 4.8. Дисциплини, включени в третия семестър на обучение	190
Таблица 4.9. Управление на риска в учебния план на специалност „Анализ и управление на данни“	192
Таблица 4.10. Разпределение на темите, включени в дисциплината „Въведение в науката за данните“	196
Таблица 4.11. Организация и структура на методологията за изучаване на „Въведение в науката за данните“	203

ИЗПОЛЗВАНА ЛИТЕРАТУРА

1. **Андреев**, А. Л. Компетентностная парадигма в образовании: опыт философско-методологического анализа. – В: *Педагогика*, 2005, № 4, с. 19 – 27.
2. **Андреев**, М. Дидактика. София: Народна просвета, 1981.
3. **Болотов**, В. А. Компетентностная модель: от идеи к образовательной программе. – В: *Педагогика*, 2003, № 10.
4. **Вацов**, С. Някои особености на компетентностния модел в немското образование. – Във: *Вісник Черкаського університету*, 2015, № 10 (343). ISSN 2076-586X.
5. **Делчева**, Ц. Система за формиране на професионални компетенции у студентите по социална педагогика. Автореферат, 2017, <http://uni-sz.bg/wp-content/uploads/truni/file/avtoreferat%20delcheva.pdf>, 20.10.2019.
6. **Дзивев**, В. Методика за обучение по модул „Бази от данни“. Дисертационен труд за присъждане на образователна и научна степен „доктор“ в професионално направление 1.3 „Педагогика на обучението по...“, научна специалност: 05.07.03 Методика на обучението по информатика и ИТ, 2017, <https://omi.fmi.uni-sofia.bg/wp-content/uploads/2018/06/VDzivevDisertaciq.pdf>
7. **Закон** за висшето образование. Обн. ДВ, бр. 112 от 27.12.1995 г., доп. бр. 86 от 18.10.2018 г., в сила от 20.01.2019 г., https://www.mon.bg/upload/17658/zkn_VO_181018.pdf
8. **Илчева**, Ю. Професионалната насоченост на обучението във ВУЗ в контекста на функциониращата в съвременната дидактика триада „знания – умения и навици – компетенции“. – В: *Годишник на Минно-геоложкия университет „Св. Иван Рилски“*, Том 51, Св. IV, Хуманитарни и стопански науки, 2008.
9. **Исаева**, Т. Е. Классификация профессионально-личностных компетенций вузовского преподавателя. – В: *Педагогика*, 2006, № 9, с. 55 – 60.
10. **Лернер**, И. Я. Формирование и развитие профессионалов в условиях современного образовательного пространства. Москва: *Издательство университета Дружбы народов*, 2006.

11. **Окуловский, О. И.** Компетенции и компетентностный подход в обучении. – В: Молодой ученый, 2012, № 12, с. 499 – 500, /<https://moluch.ru/archive/47/5841/>, 22.10.2019.
12. **Петров, П.** Дидактика. София: Веда Словена-ЖГ, 1998.
13. **Софьина, В. Н.** Развитие профессиональной компетентности специалиста в вузе и на производстве. Монография. СПб.: ГГУ им. А. С. Пушкина, 2005. 144 с.
14. **Татур, Ю. Г.** Компетентность в структуре модели качества подготовки специалиста. – Ввв: Высшее образование сегодня, 2004, № 3, с. 20 – 26.
15. **Теодосиев, Т.** Таксономията на Блум и обучението в стил на програмиране. VI Национална конференция „Образованието в информационното общество“, с. 227 – 236.
16. **Терзиев, В.** Обучение, основано на компетенции, в публичните служби за заетост (Competence-Based Training in Public Employment Services). Tenth International Scientific Conference THE POWER OF KNOWLEDGE, Tenth International Scientific Conference THE POWER OF KNOWLEDGE, 2016, с. 33 – 48, <https://ssrn.com/abstract=3158324>
17. **Толева-Стоименова, С., Д. Христозов.** Процес на информиране в уеббазираните социални мрежи. Научна конференция „Младите – инвестиция в бъдещето“, организирана от УниБИТ с подкрепата на ОПЧР по проект BG051PO001-3.3.06-0055 „Създаване на инфомационна среда за мотивация и стимулиране на младите изследователи в УниБИТ“, 18 – 19 октомври 2013 г., София.
18. **Толева-Стоименова, С., Д. Христозов.** Процеси и системи за информиране. София: За буквите – О писменехъ, 2014.
19. **Ушева, М.** Компетентностният подход – ресурс, капитал и конкурентно предимство в съвременния мениджмънт. – В: *Управление и устойчиво развитие*, 2011, № 3, с. 40 – 46.
20. **Aasheim, C., S. Williams, P. Rutner, A. Gardiner.** Data Analytics vs. Data Science: A Study of Similarities and Differences in Undergraduate Programs Based on Course Descriptions. – In: *Journal of Information Systems Education*, Spring, 2015, Vol. 26, No. 2.
21. **Aboelela, S. W., E. Larson, S. Bakken, O. Carrasquillo, A. Formicola, S. A. Glied, J. Haas, K. M. Gebbie.** Defining Interdisciplinary Research: Conclusions from a Critical Review of the Literature. – In: *Health Services Research*, 2007, Vol. 42, pp. 329 – 346.
22. **Adams, M. N.** Perspectives on Data Mining. – In: *International Journal of Market Research*, 2010, Vol. 52, No.1, pp. 11 – 19.

23. **Adelman, C., P. Ewell, P. Gaston, C. G. Schneider.** Degree Qualification Profile. Lumina Foundations Publications, 2014, <https://www.luminafoundation.org/resources/dqp>
24. **Ala-Mutka, K.** Mapping digital competence: Towards a conceptual understanding. Luxemburg, European Union, 2011.
25. **Anderson, J. W., D. R. Krathwhol, P. W. Airasia.** Taxonomy for learning, teaching and assessing: a revision of Bloom's taxonomy of education. New York: Person Education, 2003.
26. **Andreasen, N. C., T. Brown et al.** National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. Facilitating Interdisciplinary Research. Washington, DC: The National Academies Press, 2005. <https://www.nap.edu/read/11153/>
27. **Andrejevic, M.** Big Data Surveillance. – In: *Surveillance and Society*, 2014, Vol. 12, No. 2.
28. **Andrejevic, M.** Big Data, Big Questions. The Big Data Divide. – In: *International Journal Of Communication*, 2014, Vol. 8, No. 17, <https://ijoc.org/index.php/ijoc/article/view/2161/1163>
29. **Ayankoya, K., A. Calitz, J. Greyling.** Intrinsic Relations between Data Science, Big Data, Business Analytics and Datafication. Proceedings of the SAICSIT '14, 2014, pp. 192 – 198.
30. **Agarwal, S.** Understanding the Data Science Lifecycle, 2018, <http://sudeep.co/data-science/Understanding-the-Data-Science-Lifecycle/>
31. **Bamforth, C., R.** Bridging the big data divide. Part I – What your business needs to ask of its IT department. Quocirca, May 2014, <https://www.em360tech.com/wp-content/uploads/pdf/Big-Data-Divide.pdf>
32. **Bateman, S., C. Gutwin, M. Nacenta.** Seeing things in the clouds: the effect of visual features on tag cloud selections. Proceedings of the nineteenth ACM conference on Hypertext and hypermedia, 2008, pp. 193 – 202.
33. **Benjamins, V.** BD: From hype to reality? Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14), ACM. New York, NY, USA, 2014, pp. 2:1 – 2:2.
34. **Berry, M., G. Linoff.** Mastering data mining: The art and science of customer relationship management. Hoboken, NJ: John Wiley & Sons, 2001.
35. **Beyer, M., D. Laney.** The Importance of “Big Data”: A Definition. – In: **Gartner** report, 2012, pp. 1 – 9.

36. **Bharadwaj, A.** A resource-based perspective on information technology capability and firm performance: an empirical investigation. – In: *MIS quarterly*, 2000, Vol. 24, No. 1, pp. 169 – 196.
37. **Big Data Working Group, Big Data Taxonomy, Cloud Security Alliance**, 2014, https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Taxonomy.pdf
38. **Bihani, P., S. T. Patil.** A comparative study of data analysis techniques. – In: *International Journal of Emerging Trends & Technology in Computer Science*, 2014, Vol. 3, No. 2, pp. 95 – 101.
39. **Bing, L.** *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, May 2012. IBM and Said Business School. *Analytics: The real-world use of big data: How innovative enterprises extract value from uncertain data*. IBM Institute for Business Value and Said Business School Executive Report, Oct. 2012.
40. **Blockeel, H., J. Vanschoren.** Experiment databases: Towards an improved experimental methodology in machine learning. – In: **Kok, J., J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenic, A. Skowron** (eds.). *Knowledge Discovery in Databases: PKDD 2007*, Lecture Notes in Computer Science, Vol. 4702, pp. 6 – 17. Springer Berlin / Heidelberg, 2007, <http://dx.doi.org/10.1007/978-3-540-74976-9>
41. **Bloom, B., M. Engelhart, E. Furst, W. Hill, D. Krathwohl.** *Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay, 1956.
42. **Booth, A., D. Hendrix.** Libraries and Institutional Data Analytics: Challenges and Opportunities. – In: *Journal of Academic Librarianship*, 2015, Vol. 41, No. 5, pp. 695 – 699, doi: 10.1016/j.acalib.2015.08.001.
43. **Boulis, K.** What is difference between a data analyst and a data scientist?, 2014. <https://www.quora.com/What-is-difference-between-a-data-analyst-and-a-data-scientist>
44. **Boyd, D., K. Crawford.** Six provocations for big data. Presentation at A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, Oxford Internet Institute, Oxford University, Oxford, UK, 2011, <http://dx.doi.org/10.2139/ssrn.1926431>
45. **Boyd, D., K. Crawford.** Critical Questions for BD: Provocations for a Cultural, Technological, and Scholarly Phenomenon. – In: *Information, Communication, & Society*, 2012, Vol. 15, pp. 662 – 679.
46. **Brown, D., J. Kusiak.** *Problem Analysis Techniques*, IRM Training, White Paper, <https://www.miun.se/siteassets/fakulteter/nmt/summer-university/problemanalysispdf>

47. **Cao, L.** Data science: A comprehensive overview. – In: *ACM Computing Surveys*, June 2017, Vol. 50, No. 3, Article 43.
48. **Cebr:** Data equity, Unlocking the value of big data. – In: *SAS Reports*, 2012, pp. 1 – 44.
49. **Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth.** CRISP-DM 1.0: Step-by-step data mining guide. CRISP-DM consortium, 2000.
50. **Chen, H., R. Chiang, V. Storey.** Business Intelligence and Analytics: From Big Data to Big Impact. – In: *MIS quarterly*, 2012, Vol. 36, No. 4, pp. 1165 – 1188, <https://elsekuipers.wordpress.com/tag/pre-emptive-analytics/>
51. **Choo, C. W.** The Knowing Organisation: How organizations use information to construct meaning, create knowledge and make decisions. – In: *International Journal of Information Management*, 1996, Vol. 16, No. 5, pp. 329 – 340, <http://choo.fis.utoronto.ca/FIS/respub/IJIM1996.pdf>
52. **Christozov, D., S. Denchev, S. Toleva-Stoimenova, K. Rasheva-Yordanova.** Training information brokers: A curriculum model. – In: *Journal of Issues in Informing Science and Information Technology*, 2008, No. 5, pp. 87 – 94.
53. **Christozov, D., S. Toleva-Stoimenova.** The role of information brokers in knowledge management. – In: *Online Journal of Applied Knowledge Management. A Publication of the International Institute for Applied Knowledge Management*, 2014, Vol. 2, Issue 2, pp. 109 – 119, <https://pdfs.semanticscholar.org/7940/7848512eb0864224c2dd475645aa40b524f9.pdf>, 24.11.2019.
54. **Christozov, D., S. Toleva-Stoimenova, K. Rasheva-Yordanova, I. Vukarski.** Developing Big Data Competences in the Digital Era. Big data, Knowledge and Control Systems Engineering, BdKCSE'2016. pp. 97 – 104.
55. **Christozov, D., S. Toleva-Stoimenova.** Big Data Literacy – a New Dimension of Digital Divide: Barriers in learning via exploring Big Data, in *Strategic Data Based Wisdom in the Big Data Era*. Editors: J. Girard, K. Berg, D. Klein. IGI Global, 2015.
56. **Christozov, D., K. Rasheva-Yordanova, S. Toleva-Stoimenova.** Challenges in Designing Curriculum for Trans-Disciplinary Education: On Cases of Designing Concentration on Informing Science and Master Program on Data Science Informing Science. – In: *International Journal of an Emerging Transdiscipline*, 2019, Vol. 22, pp. 19 – 30.

57. **Christozov, D., S. Toleva-Stoimenova, K. Rasheva-Yordanova.** Analytical competences in big data era: taxonomy. Proceedings of ICERI2018 Conference. 12th-14th November 2018, Seville, Spain, pp. 7182 – 7191.
58. **Christozov, D., S. Toleva-Stoimenova.** Knowledge diffusion via social networks: the 21st century challenge. – In: *International Journal of Digital Literacy and Digital Competence*, 2013, Vol. 4, No. 2, April-June, pp. 1 – 12.
59. **Cleveland, W. S.** Data science: an action plan for expanding the technical areas of the field of statistics. – In: *International statistical review*, 2001, Vol. 69, No. 1, pp. 21 – 26.
60. **Columbus, L.** IBM predicts demand for data scientists will soar 28% by 2020, <https://www.forbes.com>, May 13, 2017.
61. **Coyle, P.** What Does It Mean to be a Senior Data Scientist?, 2018, <https://www.dataquest.io/blog/what-does-it-mean-to-be-a-senior-data-scientist/>
62. **Conway, D.** Data Science Venn Diagram, 2010, <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
63. **Cooper, M., P. Mell.** Tackling Big Data. Federal Computer Security Managers' Forum, 2012, http://csrc.nist.gov/groups/SMA/forum/documents/june2012presentations/fcsm_june2012_cooper_mell.pdf
64. **Costa, C., M. Santos.** A Conceptual Model for the Professional Profile of a Data Scientist. – In: **Rocha, Á., A. Correia, H. Adeli, L. Reis, S. Costanzo** (eds.). *Recent Advances in Information Systems and Technologies*. – In: *Advances in Intelligent Systems and Computing*, 2017, Vol. 570, Springer.
65. **Costa, C., M. Santos.** The Data Scientist Profile and its Representativeness in the European e-Competence Framework and the Skills Framework for the Information Age. – In: *International Journal of Information Management: The Journal for Information Professionals*, 2017, Vol. 37, Issue 6, pp. 726 – 734.
66. **Cottrel, S.** *Critical Thinking Skills Developing Effective Analysis and Argument*. Palgrave-Macmillan: Basingstoke, 2005
67. **Cottrel, S.** *The Study Skills Handbook*. 2nd Edn. Palgrave-Macmillan: Basingstoke, 2003. Retrieved from [https://elearn.unisofia.bg/pluginfile.php/76858/mod_resource/content/1/%5BStella_Cottrell%5D_Critical_Thinking_Skills_Develo\(BookFi.org\).pdf](https://elearn.unisofia.bg/pluginfile.php/76858/mod_resource/content/1/%5BStella_Cottrell%5D_Critical_Thinking_Skills_Develo(BookFi.org).pdf)
68. **Cox, M., D. Ellsworth.** Application-controlled demand paging for out-of-core visualization. Proceedings of the IEEE 8th conference on Visualization, 1997, pp. 235 – 244.

69. **Cukier**, K. N. The thing, and not the thing. – In: *The Economist*, 2013, <http://www.economist.com/blogs/graphicdetail/2013/02/elusive-big-data>
70. **Daniel**, B. Big data and analytics in higher education: Opportunities and challenges. – In: *British Journal of Education Technologies*, 2015, Vol. 46, pp. 904 – 920.
71. **Dave**, R. H. Psychomotor Levels. – In: **Developing** and Writing Behavioral Objectives, ed. Robert J. Armstrong. Tucson AZ: Educational Innovators Press, 1970.
72. **Davenport**, T. H. Analytics 3.0. – In: *Harvard Business Review*, 2013, Vol. 91, No. 12, pp. 64 – 72.
73. **Davenport**, T. H. & D. J. **Patil**. Data Scientist: The Sexiest Job of the 21st Century. – In: *Harvard Business Review*, 90, 10, October 2012, pp. 70 – 76. Retrieved February 02, 2019, from <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
74. **Davydov**, V., V. **Zinchenko**, N. **Talyzina**. The problem of activity in the works of A. N. Leontiev. – In: *Soviet Psychology*, 1982, No. 21, pp. 31 – 42.
75. **Debortoli**, S., O. **Müller**, J. vom **Brocke**. Comparing business intelligence and big data skills. – In: *Business & Information Systems Engineering*, 2014, Vol. 6, No. 5, pp. 289 – 300.
76. **Demchenko**, Y. The Emerging Role of the Data Scientist and the experience of Data Science education at the University of Amsterdam. LEARN Toolkit of Best Practice for Research Data Management. 2017, DOI: <https://doi.org/10.14324/000.learn.00>
77. **Demchenko**, Y., A. **Belloum**, T. **Wiktorski**. Edison Data Science Framework: Part 3. – In: *Data Science Model Curriculum (MC-DS)*, Release 2, 2017.
78. **Demchenko**, Y., P. **Membrey**, P. **Grosso**, C. de **Laat**. Addressing big data issues in scientific data infrastructure. Proceedings of the IEEE International Conference on Collaboration Technologies and Systems (CTS '13), 2013, pp. 48 – 55.
79. **Demchenko**, Y., C. de **Laat**, P. **Membrey**. Defining Architecture Components of the Big Data Ecosystem. Proceeding of International Conference on Collaboration Technologies and Systems (CTS'2014), Minneapolis, USA, 2014, pp. 104 – 112.
80. **Dhar**, V. Data Science and Prediction. – In: *Communication of the ACM*, 2013, Vol. 56, No. 12, <http://doi.org/10.1145/2500499>
81. **Dichev**, Ch., D. **Dicheva**. Towards Data Science Literacy. – In: *Procedia Computer Science*, 108C, 2017, pp. 2151 – 2160.

82. **Dijcks, J.** Big Data for the Enterprise. – In: **Oracle White Paper**, June, 2012.
83. **Diamantini, C., D. Potena, E. Storti.** Ontology-driven kdd process composition. – In: **IDA, 2009**, pp. 285 – 296.
84. **Diggle, P.** Statistics: A Data Science for the 21st Century. – In: *Royal Statistical Society*, 2015, 178 (Part 4), p. 20.
85. **Donoho, D.** 50 Years of Data Science. – In: *Journal of Computational and Graphical Statistics*, 2017, Vol. 26, Issue 4.
86. **Elshawi, R., S. Sakr, D. Talia, P. Trunfio.** Big data systems meet machine learning challenges: Towards big data science as a service. Big data research, 2018.
87. **Engestrom, Y.** Activity theory and individual and social transformation. – In: *Multidisciplinary Newsletter for Activity Theory*, 1991, 7/8, pp. 6 – 17.
88. **Engestrom, Y.** Activity theory and individual and social transformation. – In: **Engestrom, Y., R. Miettinen, & R.-L. Punamaki** (eds.). Perspectives on activity theory (pp. 19 – 38). Cambridge: Cambridge University Press, 1999.
89. **Engestrom, Y.** Activity theory as a framework for analyzing and redesigning work. – In: *Ergonomics*, 2000, 43(7), pp. 960 – 974.
90. **Ennis, R. H.** Critical thinking and subject specificity: Clarification and needed research. – In: *Educational Researcher*, 1989, Vol. 18, No. 3, pp. 4 – 10.
91. **Ennis, R. H.** Critical Thinking Assessment. – In: *Theory into Practice*, 1993, Vol. 32, pp. 179 – 186.
92. **Ermolayev, V., M. Suárez-Figueroa, O. Molchanovskiy.** Architecting Data Science Education. – In: **ICTERI 2018**. Vol. II: Workshops. Kyiv, Ukraine, 2018.
93. **Eshet-Alkalai, Y.** Digital literacy: A conceptual framework for survival skills in the digital era. – In: *Journal of Educational Multimedia and Hypermedia*, 2004, Vol. 13, No. 1, pp. 93 – 106.
94. **Eshet-Alkalai, Y., E. Chajut.** Changes over time in digital literacy. – In: *Cyberpsychology & Behavior*, 2009, 12(6), pp. 713 – 715.
95. **Espinosa, R., D. García-Saiz, M. Zorrilla, J. Zubcoff, J.-N. Mazón.** Enabling Non-expert Users to Apply Data Mining for Bridging the Big Data Divide. – In: *Data-Driven Process Discovery and Analysis*, 2015, Vol. 203, pp. 65 – 86.
96. **Eubanks, C.** Three Lessons CrossFit Taught Me About Data Science, 2016, <https://blogs.gartner.com/christi-eubanks/three-lessons-crossfit-taught-data-science/>

97. **Facione**, P. A. The California Critical Thinking Skills Test college level, Technical report #2. Factors predictive of CT skills. Millbrae, CA: California Academic Press, 1990.
98. **Faghmous**, J., **A. Banerjee**, **S. Shekhar**, **M. Steinbach**, **V. Kumar**, **A. R. Ganguly**, **N. Samatova**. Theory-Guided Data Science for Climate Change. – In: *Computer*, Nov. 2014, Vol. 47, No. 11, pp. 74 – 78.
99. **Fan**, W., **A. Bifet**. Mining big data: current status, and forecast to the future. – In: *ACM SIGKDD Explorations Newsletter*, 2013, Vol. 14, pp. 1 – 5.
100. **Fayyad**, U. M., **G. Piatetsky-Shapiro**, **P. Smyth**. The kdd process for extracting useful knowledge from volumes of data. – In: *Commun. ACM*, 1996, Vol. 39, No. 11, pp. 27 – 34.
101. **Firican**, G. The 10 Vs of big data, Upside, 2017, <https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx>
102. **Fink**, L. D. Creating significant learning experiences: An integrated approach to designing college courses. San Francisco: Jossey Bass, 2003.
103. **Gandomi**, A., **M. Haider**. Beyond the hype: BD concepts, methods, and analytics. – In: *International Journal of Information Management*, 2015, Vol. 35, No. 2, pp. 137 – 14.
104. **Gantz**, J., **E. Reinsel**. Extracting Value from Chaos. IDC’s Digital Universe Study, 2011.
105. **Gehl**, R. W. Sharing, knowledge management and big data: A partial genealogy of the data scientist. – In: *European Journal of Cultural Studies*, 2015, Vol. 18 (4 – 15), pp. 413 – 428, <http://doi.org/10.1177/1367549415577385>
106. **Geringer**, S. Data Science Venn Diagram v2.0 Retrieved, 2014, <http://www.anlytcs.com/2014/01/data-science-venn-diagram-v20.html>
107. **Girard**, J., **D. Klein**, **K. Berg** (eds.). Strategic Data-Based Wisdom in the Big Data Era. Hershey, PA: IGI Global, 2015.
108. **Glister**, P. Digital Literacy. New York, Wiley, 1997.
109. **Granville**, V. Developing Analytic Talent: Becoming a Data Scientist. John Wiley and Sons, Incorporated, US, 2014.
110. **Granville**, V. 22 Differences Between Junior and Senior Data Scientists, 2018, <https://www.datasciencecentral.com/profiles/blogs/10-differences-between-junior-and-senior-data-scientist>
111. **Grus**, J. Post-Prism Data Science Venn Diagram, <http://joelgrus.com/2013/06/09/post-prism-data-science-venn-diagram/>

112. **Guilford**, J. P. Rotation problems in factor analysis. – In: *Psychological Bulletin*, 1974, Vol. 81, pp. 498 – 501.
113. **Guilford**, J. P. Factors and factors of personality. – In: *Psychological Bulletin*, 1975, Vol. 82, pp. 802 – 814.
114. **Guilford**, J. P., **R. Hoepfner**. The Analysis of Intelligence. New York: McGraw-Hill, 1971.
115. **Hardin**, J., **R. Hoerl**, **N. J. Horton**, **D. Nolan**, **B. Baumer**, **O. Hall-Holt**, **P. Murrell**, **R. Peng**, **P. Roback**, **D. Temple Lang**, **M. D. Ward**. Data science in statistics curricula: Preparing students to “think with data”. – In: *The American Statistician*, 2015, Vol. 69, pp. 343 – 353.
116. **Harris**, H. The Data Products Venn Diagram, 2013, <http://www.datacommunitydc.org/blog/2013/09/the-data-products-venn-diagram>
117. **Harris**, J. G., **N. Shetterley**, **A. E. Alter**, **K. Schnell**. The Team Solution to the Data Scientist Shortage. Accenture Institute for High Performance, 2013, [http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture e-Team-Solution-Data-ScientistShortage.pdf](http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-e-Team-Solution-Data-ScientistShortage.pdf)
118. **Harrington**, L. Clinical intelligence. – In: *Journal of Nursing Administration*, 2011, Vol. 41, No. 12, pp. 507 – 509.
119. **He**, Y., **R. Lee**, **Y. Huai**, **Z. Shao**, **N. Jain**, **X. Zhang**, **Z. Xu**. RCFfile: A Fast and Spaceefficient Data Placement Structure in MapReduce-based Warehouse Systems. IEEE International Conference on Data Engineering (ICDE), 2011, pp. 1199 – 1208.
120. **Heerschop**, B., **A. Hogenboom**, **F. Frasinca**. Sentiment lexicon creation from lexical resources. Business Information Systems, 2011.
121. **Herodotou**, H., **H. Lim**, **G. Luo**, **N. Borisov**, **L. Dong**, **F. B. Cetin**, **S. Babu**. Starfish: A Self-tuning System for Big Data Analytics. – In: **Proceedings** of the Conference on Innovative Data Systems Research, 2011, pp. 261 – 272.
122. **Hicks**, S. C., **R. A. Irizarry**. A guide to teaching Data Science. arXiv:1612.07140, 2017.
123. **Hilario**, M., **A. Kalousis**, **P. Nguyen**, **A. Woznica**. A data mining ontology for algorithm selection and meta-mining. ECML/PKDD09 Workshop on Third Generation Data Mining: Towards ServiceOriented Knowledge Discovery. 2009, pp. 76 – 87.
124. **Hilario**, M.: e-lico annual report 2010. Tech. rep., Université de Geneve, 2010.
125. **Ho**, T. C., **S. C. Mat**, **L. H. San** et al. A prediction model for co2 emission from manufacturing industry and construction in malaysia.

- International Conference on Space Science and Communication (IconSpace), IEEE, 2015, pp. 469 – 472.
126. **How** to Practice Your Data Science Skills
<https://www.kdnuggets.com/2018/05/simplilearn-9-must-have-skills-data-scientist.html>
 127. **Hu, M., B. Liu.** Mining and summarizing customer reviews. – In: **Proceedings** of KDD, 2004.
 128. **IBM.** Bringing big data to the enterprise, 2012.
<http://www01.ibm.com/software/in/data/bigdata>
 129. **Intel.** Big Data Analytics. Intel’s IT Manager Survey on How Organizations Are Using Big Data, 2012.
 130. **Ismail, N.** Data scientist skills. – In: *IOSR Journal of Mobile Computing & Application (IOSR – JMCA)*, Jul.-Aug. 2016, Vol. 3, Issue 4, pp. 52 – 61.
 131. **Joseph, R. C., N. A. Johnson.** Big data and transformational government. – In: *IT Professional*, 2013, Vol. 15, No. 6, pp. 43 – 48.
 132. **Kaisler, Stephen et al.** BD: Issues and challenges moving forward. – In: **Proceedings** of 46th Hawaii International Conference on System Sciences (HICSS), Hawaii, 2013, pp. 995 – 1004.
 133. **Kameswari, U. S., I. R. Babu.** Sensor data analysis and anomaly detection using predictive analytics for process industries. IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCID), 2015, pp. 1 – 8.
 134. **Kalechofsky, H.** A Little Data Science Business Guide, 2016,
<http://www.msquared.com/wp-content/uploads/2017/01/A-Simple-Framework-for-Building-Predictive-Models.pdf>
 135. **Kietz, J. U., F. Serban, A. Bernstein, S. Fischer.** Designing kdd-workflows via htn-planning. – In: **Raedt, L. D., C. Bessière, D. Dubois, P. Doherty, P. Frasconi, F. Heintz, P. J. F. Lucas** (eds.). ECAI. Frontiers in Artificial Intelligence and Applications, IOS Press, 2012, Vol. 242, pp. 1011 – 1012.
 136. **Kowalczyk, M., P. Buxmann.** Big Data and information processing in organizational decision processes. – In: *Business & Information Systems Engineering*, 2014, Vol. 6, No. 5, pp. 267 – 278.
 137. **Kriegel, H. P., K. M. Borgwardt, P. Kröger, A. Pryakhin, M. Schubert, A. Zimek.** Future trends in data mining. – In: *Data Min. Knowl. Discov.*, 2007, Vol. 15, No. 1, pp. 87 – 97.
 138. **Ku, K.** Assessing students’ critical thinking performance: Urging for measurements using multiresponse format. – In: *Thinking Skills and Creativity*, 2009, Vol. 4, pp. 70 – 76.

139. **Laney, D.** 3D data management: Controlling data volume, velocity and variety. – In: *META Group Research Note*, February 2001, No. 6.
140. **Larson, D., V. Chang.** A review and future direction of agile, business intelligence, analytics and data science. – In: *International Journal of Information Management*, 2016, Vol. 36, No. 5, pp. 700 – 710.
141. **LaValle, S., E. Lesser, R. Shockley, M. S. Hopkins & N. Kruschwitz.** Big data, analytics and the path from insights to value. – In: *MIT sloan management review*, 52 (2), 21, 2011.
142. **Leontiev, A. N.** Activity, consciousness, and personality (originally published in Russian in 1975). Prentice-Hallp, 1978.
143. **Leu, D., L. Zawilinski, J. Castlek, M. Banerjee, B. Housand, Y. Liu, N. O’Neil.** What is new about the New Literacies of online reading comprehension? – In: L. **Rush, A. Eakle & A. Berger** (eds.). Secondary school literacy: What research reveals for classroom. Urbana, IL, National Council of Teachers of English, 2007.
144. **Liang, H., J. Zhu.** Big Data, Collection of (Social Media, Harvesting). In: **The International Encyclopedia of Communication Research Methods**, 2017, http://weblab.com.cityu.edu.hk/blog/wp-content/uploads/2017/12/Liang_Zhu_2017_Big-data.pdf
145. **Liñán, L., J. Pérez.** Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. – In: *RUSC. Universities and Knowledge Society Journal*, 2015, Vol. 12, No. 3, pp. 98 – 112.
146. **Liu Ou, L., L. Frankel, K. C. Roohr.** Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment, Research Report, June 2014, <https://files.eric.ed.gov/fulltext/EJ1109287.pdf>Longbottom
147. **Lotrecchiano, G. R., S. Misra.** Transdisciplinary knowledge producing teams: Towards a complex systems perspective. – In: *Informing Science: The International Journal of an Emerging Transdiscipline*, 2018, Vol. 21, pp. 51 – 74, <https://doi.org/10.28945/4086>
148. **Loukides, M.** What is data science. An O’Reilly Radar Report, 2010, <http://radar.oreilly.com/2010/06/what-is-data-science.html>
149. **Lyon, D.** (ed.). Surveillance as social sorting: Privacy, risk and automated discrimination. New York, NY: Routledge, 2002.
150. **Maasen, S., O. Lieven.** Transdisciplinarity: a new mode of governing science? – In: *Science & Public Policy*, 2006, Vol. 33, No. 6, pp. 399 – 410.
151. **Manieri, A., Y. Demchenko, S. Brewer, M. Hemmje, R. Riestra & J. Frey.** Data Science Professional uncovered How the EDISON Project will contribute to a widely accepted profile for Data Scientists. Paper

- presented at the IEEE 7th International Conference on Cloud Computing Technology and Science Data, 2015.
152. **Mantri**, S. Data Science: Literature Review & State of Art, 2016.
 153. **Manyika**, J., **M. Chui**, **B. Brown**, **J. Bughin**, **R. Dobbs**, **C. Roxburgh**, **A. Byers**. *BD: The next frontier for innovation, competition, and productivity*. McKinsey & Company, 2011.
 154. Markow, W., Braganza, S., Taska, B. et al. The Quant Crunch How The Demand For Data Science Skills Is Disrupting The Job Market. Burning Glass Technologies 2017. IBM //https://www.ibm.com/downloads/cas/3RL3VXGA.
 155. **Marwick**, A., **D. Boyd**. To See and Be Seen: Celebrity Practice on Twitter. – In: *Convergence: The International Journal of Research into New Media Technologies*, 2011, vol. 17, pp. 139 – 158.
 156. **Matter**, U. Data Science in Business/Computational Social Science in Academia? 2013, <http://giventheedata.blogspot.com/2013/03/data-science-in-businesscomputational>
 157. **Mayer-Schoenberger**, V., **K. Cukier**. *Big Data. A Revolution that will transform how we live, work, and think*. London: John Murray Publishers, 2013.
 158. **Metiri** Group – Commissioned by Cisco. *Multimodal Learning Through Media: What the Research Says*. Cisco Systems, Inc., 2008.
 159. **Microsoft**. *The Big Bang: How the Big Data Explosion Is Changing the World*, 2013.
 160. **Microsoft** Azure Blog, <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/lifecycle-modeling>
 161. **Microsoft** Azure Blog, <https://azure.microsoft.com/en-us/blog/?v=18.37>
 162. **Mikalef**, P., **I. Pappas**, **M. Giannakos** & **J. Krogstie**. The Human Side of Big Data Understanding the skills of the data scientist in education and industry. Paper presented at the IEEE EDUCON 2018 Global Engineering Education Conference, Tenerife, Canary Islands, Spain, 2018.
 163. **Mikalef**, P., **I. Pappas**, **J. Krogstie**, **M. Giannakos**. Big data analytics capabilities: a systematic literature review and research agenda. – In: **Information Systems and e-Business Management**, 2017, pp. 1 – 32.
 164. **Molina-Solana**, M., **M. Ros**, **M. D. Ruiz**, **J. Gómez-Romero**, **M. Martín-Bautista**. Data science for building energy management: A review. – In: *Renewable and Sustainable Energy Reviews*, 2017, Vol. 70, pp. 598 – 609.

165. **Marzano, R. J.** Designing a new taxonomy of educational objectives. Thousand Oaks, CA: Corwin Press, 2000.
166. **Naur, P.** Concise Survey of Computer Methods. Petrocelli Books, 1st edition, 1974.
167. **Newman, R., V. Chang, R. J. Walters, G. B. Wills.** Model and experimental development for business data science. – In: *International Journal of Information Management*, 2016, Vol. 36, No. 4, pp. 607 – 617.
168. **Nikolov, R.** The Global Campus: ICT and the Global Transformation of Higher Education. – In: *Serdica Journal of Computing*, 2010, Vol. 4, No. 2, pp. 183 – 216.
169. **Nikolov, R., E. Shoikova, E. Kovatcheva.** Competence Based Framework for Curriculum Development. 10.13140/RG.2.1.4201.6086, 2014.
170. **Nikolov, R.** Web 2.0 and the Global Transformation of Education. Invited paper. – In: **Proceedings** of the 40th Conference of the Union of Bulgarian Mathematicians, Borovets, 5 – 9 April, 2011, pp. 109 – 119.
171. **Nolan, D., T. Speed.** Stat Labs: Mathematical Statistics Through Applications. New York: Springer-Verlag, 2000.
172. **Nolan, D., D. Temple Lang.** Data Science Case Studies in R: A Case Studies Approach to Computational Reasoning and Problem Solving. Chapman and Hall/CRC, 2015.
173. **Ottenbacher, K. J., J. E. Graham, S. R. Fisher.** Data science in physical medicine and rehabilitation: Opportunities and challenges. Physical Medicine and Rehabilitation Clinics of North America, 2019.
174. **Owais, S., N. Hussein.** Extract Five Categories CPIVW from the 9 V's Characteristics of the Big Data. – In: *International Journal of Advanced Computer Science and Applications*, 2016, Vol. 7, No. 3.
175. **Palmer, S.** Data Driven Thinking A collection of essays on data-driven decision making, 2015, <https://www.shellypalmer.com/wp-content/images/2017/09/Data-DrivenThinkingShellyPalmer.pdf>
176. **Panov, P., L. N. Soldatova, S. Dzeroski.** Towards an ontology of data mining investigations. – In: *Discovery Science*, 2009, pp. 257 – 271.
177. **Parks, D., M. Dana.** Defining Data Science and Data Scientist. Graduate Theses and Dissertations, 2017, <http://scholarcommons.usf.edu/etd/7014>
178. **Parry, S.** The quest for competencies: competency studies can help you make HR decision, but the results are only as good as the study. Training.33.P., 1996, 48 p.

179. **Paul, R., L. Elder.** A Guide For Educators to Critical Thinking Competency Standarts. Standards, Principles, Performance, Indicators, and Outcomes With a Critical Thinking Master Rubric. Foundation for Critical Thinking Press, 2005, <http://www.criticalthinking.org/resources/PDF/CT-competencies%202005.pdf>
180. **Perumal, S.** Data scientist, 2015, <http://www.slideshare.net/SevugaPerumal1/a-free-orientation-on-statistical-dataanalysis-is-conducted-on-saturday-25072015-at-10-am-and-it-has-2-hours-duration>
181. **Piaget, J.** The epistemology of interdisciplinary relationships. Paper presented at the Center 944 for Educational Research and Innovation (CERI), Paris, France, 1972, <https://www.luminafoundation.org/files/resources/dqp.pdf>
182. **Piatetsky, G. R,** Python Duel As Top Analytics, Data Science software – KDnuggets 2016 Software Poll Results, <https://www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html>
183. **Powers, D. E., M. K. Enright.** Analytical reasoning skills involved in graduate study: perceptions of faculty in six fields, GRE Board Professional Report GREB No. 83-23P, December 1986, Educational Testing Service Retrieved from <https://www.ets.org/Media/Research/pdf/RR-86-43-Powers.pdf>
184. Pre-emptive analytics. MAPAD: the new way of thinking, // <https://elsekuipers.wordpress.com/tag/pre-emptive-analytics/> Posted on August 31, 2014, 02.12.2019.
185. **Press, G.** A Very Short History Of Data Science, Forbes, 2013, <http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-datascience/>
186. **Price, E., C. De Leone, N. Lasry.** Comparing educational tools using activity theory: Clickers and flashcards. – In: **AIP** Conference Proceedings, AIP, 2010, Vol. 1289, No. 1, pp. 265 – 268.
187. **Provost, F., T. Fawcett,** Data Science Its Relationship Data-Driven Decision Making, Vol.1, No.1, pp.51–59, 2013. <http://doi.org/10.1089/big.2013.1508>
188. **Provost, F., T. Fawcett.** Data Science for Business: What you need to know about data mining and data-analytic thinking. O’Reilly Media, Inc., 2013.
189. **PwC.** What’s next for the data science and analytics job market?, 2015, <https://pwc.to/2FL8GEG>
190. **Rabhi, F., M. Bandara, A. Namvar, O. Demirors.** Big Data Analytics Has Little to Do with Analytics. – In: **Beheshti, A., M. Hashmi, H.**

- Dong, W. Zhang** (eds). Service Research and Innovation. ASSRI 2015, ASSRI 2017. Lecture Notes in Business Information Processing, Vol. 234, Springer, Cham, 2018.
191. **Raeithel, A.** Semiotic self-regularization and work: An activity theoretical foundation of design. – In: **Floyd, R.** et al. Software Development and Reality Construction, Springer Verlag, 1991.
 192. **Rasheva-Yordanova, K., E. Iliev, B. Nikolova.** Analytical Thinking As A Key Competence For Overcoming The Data Science Divide. – In: Proceedings of EDULEARN18 Conference 2 – 4 July 2018, Palma, Mallorca, Spain, pp. 7892 – 7898, 2018.
 193. **Ravi, K., V. Ravi.** A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems, 2015, Vol. 89, pp. 14 – 46.
 194. **Rasheva-Yordanova, K., V. Chantov, I. Kostadinova, E. Iliev, P. Petrova, B. Nikolova.** Forming of Data Science Competence for Bridging the Digital Divide. – In: Proceedings of 9th International Conference The Future of Education, 2019. ISBN 978-88-85813-45-8, pp. 174 – 179.
 195. **Rasheva-Yordanova, K., S. Toleva-Stoimenova, D. Christozov, I. Kostadinova.** Road Map in Developing Data Science Competences. – In: Proceedings of INTED2019 Conference 11th – 13th March 2019, Valencia, Spain, ISBN: 978-84-09-08619-1, pp. 6643 – 6650.
 196. **Resnick, L.** Education and learning to think. Washington, National Academy Press, 1987;
 197. **Rioux, J.-Y., J. Baer, P. Duplessis, J. Quah, S. (Y.) Zhan.** Predictive modelling. Turning Big Data into Big Opportunities, 2018, <https://www.cia-ica.ca/docs/default-source/2018/218081e.pdf>
 198. **Robbins, J. K.** Problem solving, reasoning, and analytical thinking in a classroom environment. – In: *The Behavior Analyst Today*, 2011, Vol. 12, No.1, pp. 40 – 47.
 199. **Romero, C., S. Ventura.** Educational Data Mining: A Survey from 1995 to 2005. – In: *Expert Systems with Applications*, 2007, Vol. 33, pp. 135 – 146, 2007.
 200. **Russom, P.** Big Data Analytics. – In: **TDWI Best Practices Report**, 2011, pp. 1 – 40.
 201. **Schroeck, M., R. Shockley, J. Smart, D. Romero-Morales, P. Tufano.** Analytics: The Real-World Use of Big Data. IBM report, 2012.
 202. **Schoenherr, T., C. Speier, Pero.** Data science, predictive analytics, and big data in supply chain management: Current state and future potential. – In: *Journal of Business Logistics*, 2015, Vol. 36, No. 1, pp. 120 – 132.

203. **Shum, S. B., W. Hall, M. Keynes, R. S. J. Baker, J. T. Behrens, M. Hawksey, N. Jeffery.** Educational Data Scientists: A Scarce Breed, 2013, <http://simon.buckinghamshum.net/wp-content/uploads/2013/03/LAK13PanelEducDataScientists.pdf>
204. **Shafer, T.** The 42 V's of Big Data and Data Science, 2017, <https://www.elderresearch.com/blog/42-v-of-big-data>
205. **Shen, Z., J. Wei, N. Sundaresan, K. L. Ma.** Visual Analysis of Massive Web Session Data. – In: **Large Data Analysis and Visualization (LDAV)**, 2012, pp. 65 – 72.
206. **Sicular, S.** Big Data Analytics Failures and How to Prevent Them. Gartner, 2015.
207. **Silver, A.** The Essential Data Science Venn Diagrams, 2017, <https://towardsdatascience.com/the-essential-data-science-venn-diagram-35800c3bef40>
208. **Sivarajah, U., M. Kamal, Z. Irani, V. Weerakkody.** Critical analysis of Big Data challenges and analytical methods. – In: *Journal of Business Research*, 2017, Vol. 70, pp. 263 – 286, 2017, <https://doi.org/10.1016/j.eswa.2006.04.005>
209. **Song, I-Y., Y. Zhu.** Big Data and Data Science: what should we teach? – In: *Expert Systems, Special Issue: Big Data Trends: Modelling Management and Visualization*, 2016, Vol. 33, Issue 4, pp. 364 – 373.
210. **Soumendra Mohanty, M. J., H. S.** Big Data Imperatives Enterprise Big Data Warehouse, BI Implementations and Analytics. Apress, 2013.
211. **Spencer, L., S. Spencer.** Competence at work: models for superior performance. New York. John Wiley, 1993.
212. **Sternberg, R. J., E. L. Grigorenko.** Styles of thinking in the school. – In: *European Journal for high ability*, 1995, Vol. 6, No. 2, pp. 201 – 219.
213. **Sternberg, R. J.** A propulsion model of types of creative contributions. – In: *Review of General Psychology*, 1999, Vol. 3, pp. 83 – 100.
214. **Stockinger, K., T. Stadelmann, A. Ruckstuhl.** Data Scientist als Beruf. Big Data – Grundlagen, Systeme und Nutzungspotenziale, Springer Verlag, Edition HMD, 2015, pp. 59 – 81.
215. **Szongott, C., B. Henne, G. von Voigt.** Big data privacy issues in public social media. 6th IEEE international conference on digital ecosystems technologies (DEST), 2012, pp. 1 – 6.
216. **Taboada, M., J. Brooke, M. Tofiloski, K. Voll, M. Stede.** Lexicon-based methods for sentiment analysis. Computational Linguistics, 2011.
217. **Tang, K. N.** Development of Analytical Thinking Skills Among Thai University Students – In: *TOJET: The Turkish Online Journal of Educational Technology, Special Issue for INTE*, 2017, pp. 862 – 869.

218. **Tang, R., W. Sae-Lim.** Data science programs in U.S. higher education: An exploratory content analysis of program description, curriculum structure, and course focus. – In: *Education for Information*, Vol. 32, No. 3, 2016, pp. 269 – 290.
219. **The United Kingdom Information Commissioner’s Office.** BD, artificial intelligence, mashine learning and data protection. Version 2.2., March 2017.
220. **Tierney, B.** Data Science Is Multidisciplinary. Oralytics, June 2012 // <https://oralytics.com/2012/06/13/data-science-is-multidisciplinary/>
221. **Toleva-Stoimenova, Stefka, Katia Rasheva-Yordanova, Dimitar Christozov.** New dimensions of data science professional skills as emerged by identified ethical issues: GDPR. – In: Proceedings of ICERI2018 Conference 12th – 14th November 2018, Seville, Spain, 2018, pp. 0488 – 0497.
222. **Tukey, J. W.** Exploratory Data Analysis. Reading, MA: Addison-Wesley, 1977.
223. **Turkay, C., N. Pezzotti, C. Binnig, H. Strobel, B. Hammer, D. Keim, & J.-D. Fekete, T. Palpanas, Y. Wang, F. Rusu.** Progressive Data Science: Potential and Challenges, 2019, <https://hal.inria.fr/hal-01961871/document>
224. **van der Aalst, W. M. P.** Data Scientist: The Engineer of the Future. Proceedings of the I-ESA Conference (K. Mertins, F. Benaben, R. Poler, and J. Bourrieres, eds.), 2014, Vol. 7 of Enterprise Interoperability, Springer-Verlag, Berlin, pp. 13 – 28.
225. **Valchanov, I.** What are the Skills Needed to Become a Data Scientist in 2018?, <https://towardsdatascience.com/what-are-the-skills-needed-to-become-a-data-scientist-in-2018-d037012f1db2>
226. **Vanschoren, J., L. Soldatova.** Exposé: An ontology for data mining experiments. – In: **International Workshop on Third Generation Data Mining: Towards Service-oriented Knowledge Discovery (SoKD2010)**, 2010, pp. 31 – 46.
227. **Veaux, R.** et al. Curriculum Guidelines for Undergraduate Programs in Data Science. – In: *Annual Review of Statistics and Its Applications*, 2017, 4:2.1–2.16.
228. **Viaene, S.** Data Scientists Aren’t Domain Experts. – In: *IEEE IT Professional*, 2013, Vol. 15, Issue 6, pp. 12 – 17.
229. **Vojkovic, Goran.** Will the GDPR slow down development of Smart Cities? – In: **Proceedings of 41st International Convention on Information and Communication Technology, Electronics and**

- Microelectronics (MIPRO 2018), Opatija, Croatia, 2018, pp. 1495 – 1497.
230. **Wagner, R.** Organisational competence in project management – new perspectives on assessing and developing organisations. – In: *Journal of Project, Program & Portfolio Management*, 2012, Vol. 3, No. 1, pp. 45 – 57.
 231. **Waller, M. A., S. Fawcett.** Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. – In: *Journal of Business Logistics*, 2013, Vol. 34, No. 2, pp.77 – 84.
 232. **Wall Street Journal.** 2014. CIO's Growing Demand for Data Scientists.
 233. **Wamba, S. F., A. Gunasekaran, S. Akter, S. J.-f. Ren, R. Dubey, S. J. Childe.** Big data analytics and firm performance: Effects of dynamic capabilities. – In: *Journal of Business Research*, 2017, Vol. 70, pp. 356 – 365.
 234. **Watson, G., E. M. Glaser, I. I. Watson-Glaser.** Critical Thinking Appraisal, Technical Manual and User's Guide Pearson, 2010, https://us.talentlens.com/wp-content/uploads/wg2_techman_2012.pdf
 235. **Weinberger, D.** Too big to know: Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room. New York, Basic Books, 2011.
 236. **Wiktorski, T., Y. Demchenko, A. Belloum.** Model Curricula for Data Science EDISON Data Science Framework. 2017 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), Hong Kong, China, 2017.
 237. **Willingham, D.** Critical thinking. Why is it so Hard to Teach? – In: *American Educator*, 2007, pp. 8 – 19, https://www.aft.org/sites/default/files/periodicals/Crit_Thinking.pdf
 238. **Yan, D., G. Davis.** A First Course in Data Science. – In: *Journal of Statistics Education*, 2019, Vol. 27, Issue 2, pp. 99 – 109.
 239. **Younas, M.** Transactional services for NoSQL big data systems. – In: **Keynote** talk at the 6th international conference on multimedia computing and systems (ICMCS 2018), Rabat, Morocco, 10 – 12 May 2018.
 240. **Younas, M.** Research challenges of big data. – In: *Service Oriented Computing and Applications*, 2019, Vol. 13, Issue 2, pp. 105 – 107.
 241. **Zakova, M., P. Kremen, F. Zelezny, N. Lavrac.** Automating knowledge discovery workflow composition through ontology-based planning. – In: *IEEE T. Automation Science and Engineering*, 2011, Vol. 8, No. 2, pp. 253 – 264.

242. **Zhang, J., A. Fu, H. Wang, S. Yin.** The development of data science education in China from the LIS perspective. – In: *International Journal of Librarianship*, 2017, Vol. 2, No. 2, pp. 3 – 17.
243. **Zhang, L., A. Stoffel, M. Behrisch, S. Mittelstadt, T. Schreck, R. Pompl, S. Weber, H. Last, D. Keim.** Visual Analytics for the Big Data Era – A Comparative Review of State-of-the-Art Commercial Systems. IEEE Conference on Visual Analytics Science and Technology (VAST), 2012, pp. 173 – 182.
244. **Hayes, B.** Doing Data Science, A Book Review. Notices of the American Mathematical Society, 2014, 61(09), 1068. doi: 10.1090/noti1167

**ПУБЛИКАЦИИ, РЕАЛИЗИРАНИ
ПО ПРОЕКТ ДМ 12/04 ОТ 20.12.2017 г.,
ИЗПОЛЗВАНИ В МОНОГРАФИЯТА**

1. **Rasheva-Yordanova, K., E. Iliev, B. Nikolova.** “ANALYTICAL THINKING AS A KEY COMPETENCE FOR OVERCOMING THE DATA SCIENCE DIVIDE”. 10th annual International Conference on Education and New Learning Technologies. 2nd – 4th of July, 2018, Palma de Mallorca (Spain). IATED, 2018. (Web of Science)
2. **Iliev, E., K. Rasheva-Yordanova, V. Chantov.** “ANALYSIS OF MISSING DATA SCIENCE COMPETENCE IN IT SECTOR”. 10th annual International Conference on Education and New Learning Technologies. 2nd – 4th of July, 2018, Palma de Mallorca (Spain). IATED, 2018. (Web of Science)
3. **Christozov, D., K. Rasheva-Yordanova, S. Toleva-Stoimenova.** RISKS MANAGEMENT IN DATA SCIENCE TRAINING. – In: **Proceedings** of Regional international conference on APPLIED PROTECTION AND ITS TRENDS, Zlatibor, 7 – 10 October, 2018.
4. **Toleva-Stoimenova, S., K. Rasheva-Yordanova, D. Christozov.** NEW DIMENSIONS OF DATA SCIENCE PROFESSIONAL SKILLS AS EMERGED BY IDENTIFIED ETHICAL ISSUES: GDPR. – In: **Proceedings** of ICERI2018 Conference 12th – 14th November 2018, Seville, Spain, pp. 0488 – 0497. ISBN: 978-84-09-05948-5. (Web of Science)
5. **Christozov, D., S. Toleva-Stoimenova, K. Rasheva-Yordanova.** ANALYTICAL COMPETENCES IN BIG DATA ERA: TAXONOMY. – In: **Proceedings** of ICERI2018 Conference. 12th-14th November 2018, Seville, Spain. pp. 7182-7191, ISBN: 978-84-09-05948-5. (Web of Science)
6. **Christozov, D., K. Rasheva-Yordanova, S. Toleva-Stoimenova.** DESIGNING DATA SCIENCE CURRICULUM IN A WAY TO ADDRESS EXPECTED STUDENTS ENTRY COMPETENCES. – In: **Proceedings** of INTED2019 Conference 11th – 13th March 2019, Valencia, Spain, pp. 2635 – 2640, ISBN: 978-84-09-08619-1.
7. **Rasheva-Yordanova, K., S. Toleva-Stoimenova, D. Christozov, I. Kostadinova.** ROAD MAP IN DEVELOPING DATA SCIENCE

- COMPETENCES. – In: **Proceedings** of INTED2019 Conference 11th – 13th March 2019, Valencia, Spain, pp. 6643 – 6650, ISBN: 978-84-09-08619-1.
8. **Toleva-Stoimenova, S., D. Christozov, K. Rasheva-Yordanova.** ENTRY COMPETENCES ASSESSMENT OF DATA SCIENCE POTENTIAL STUDENTS. – In: **Proceedings** of INTED2019 Conference 11th – 13th March 2019, Valencia, Spain, pp. 4248 – 4256, ISBN: 978-84-09-08619-1.
 9. **Christozov, D., K. Rasheva-Yordanova, S. Toleva-Stoimenova.** CHALLENGES IN DESIGNING CURRICULUM FOR TRANS-DISCIPLINARY EDUCATION: ON CASES OF DESIGNING CONCENTRATION ON INFORMING SCIENCE AND MASTER PROGRAM ON DATA SCIENCE. – In: *Informing Science: The International Journal of an Emerging Transdiscipline*, 2019, Vol. 22, pp. 019 – 030, <https://doi.org/10.28945/4300>
 10. **Toleva-Stoimenova, S., D. Christozov, K. Rasheva-Yordanova.** INTRODUCTION OF EMERGING TECHNOLOGY INTO HIGHER EDUCATION CURRICULUM: THE CASE OF BLOCKCHAIN TECHNOLOGY AS PART OF DATA SCIENCE MASTER PROGRAM. – In: **Proceedings** of 9th International Conference The Future of Education, 2019, pp. 186 – 190, ISBN 978-88-85813-45-8.
 11. **Rasheva-Yordanova, K., V. Chantov, I. Kostadinova, E. Iliev, P. Petrova, B. Nikolova.** FORMING OF DATA SCIENCE COMPETENCE FOR BRIDGING THE DIGITAL DIVIDE. – In: **Proceedings** of 9th International Conference The Future of Education, 2019, pp. 174 – 179, ISBN 978-88-85813-45-8.
 12. **Rasheva-Yordanova, K., S. Toleva-Stoimenova, D. Christozov, I. Kostadinova, B. Nikolova.** ASSESSING THE ENTRY COMPETENCES OF DATA SCIENCE MASTER PROGRAM POTENTIAL STUDENTS. – In: **Proceedings** of EDULEARN19 Conference 1st – 3rd July 2019, Palma, Mallorca, Spain, pp. 4389 – 4394, ISBN: 978-84-09-12031-4.
 13. **Rasheva-Yordanova, K., S. Toleva-Stoimenova, D. Christozov,** DATA SCIENCE: CHALLENGES AND TRENDS. – In: **Proceedings** of ICERI 2019 Conference, 11-13 November, 2019, Seville, Spain, ISBN: 978-84-09-14755-7

РЕЦЕНЗИЯ

НА МОНОГРАФИЯ „ИЗГРАЖДАНЕ НА DATA SCIENCE КОМПЕТЕНТНОСТ ЗА ПРЕОДОЛЯВАНЕ НА ДИГИТАЛНОТО НЕРАВЕНСТВО“

с автори д-р Катя Рашева-Йорданова
и д-р Стефка Толева-Стоименова

**от доц. д-р Волин Карагъзов,
почетен професор на Американски университет
в България**

Темата на монографията безспорно е много актуална. Развитие-то на съвременните информационни технологии достигна етап, в който данните и уменията да се извлича полезна информация и знания от тях определят успеха или провала на членовете, физически и юридически, на обществото. Наблюдава се и нов етап в „дигиталното разделение“, характерен с уменията да се извлича стойност от натрупаните големи масиви от хетерогенни данни. Това развитие е и предизвикателство пред образователните институции, които трябва да отговорят на потребностите на обществото от компетентни специалисти по данни.

Монографията е развита в четири глави, въведение, заключение, използвани източници и приложения. Първата глава въвежда основните понятия и характеристики на областта от знания определяна като Data Science. Във втората е направен преглед на компетентностите, както според търсенето от индустрията, така и описани в литературни източници. Направена е класификация на компетентностите, които един специалист по данни трябва да притежава. В трета глава е предложен образователен модел основан на компетенции, ориентиран към изграждане на специалисти в областта на Data Science на магистърско ниво. Взето е под внимание подготовеността на студентите от различни бакалавърски специалности. Направен е извода, че специалността от бакалавърското обучение не е ключово за успешна кари-

ера като специалист по данни, но играе важна психологическа роля при взимане на решение за насочване към такава кариера и съответно кандидатстване в магистърски програми в областта. В четвърта глава е предложена структура на магистърска програма по Data Science - на български „Управление и анализ на данни“.

Монографията е издържана от методологична гледна точка и представя завършен изследователски и развоен проект. Изчерпателният преглед на литературни източници, емпиричните изследвания, като анкети със студенти и изследване на обявите за работа, дават увереност в правилността на направените изводи и обективността на базата, върху която е изградена учебната програма.

Като се има предвид положения труд и обема на извършената работа и получените резултати предлагам монографията **ИЗГРАЖДАНЕ НА DATA SCIENCE КОМПЕТЕНТНОСТ ЗА ПРЕОДОЛЯВАНЕ НА ДИГИТАЛНОТО НЕРАВЕНСТВО** да бъде издадена.

13.11.2019
София

Доц. д-р Волин Карагъзов

РЕЦЕНЗИЯ

НА МОНОГРАФИЯ „ИЗГРАЖДАНЕ НА DATA SCIENCE КОМПЕТЕНТНОСТ ЗА ПРЕОДОЛЯВАНЕ НА ДИГИТАЛНОТО НЕРАВЕНСТВО“

с автори д-р Катя Рашева-Йорданова
и д-р Стефка Толева-Стоименова

от доц. д-р **Евгения Ковачева**

Монографията „Изграждане на data science компетентност за преодоляване на дигиталното неравенство“ има амбициозната задача да изследва компетенциите, които новото поколение експерти трябва да притежават – експертите в областта на *науката за данните и технологиите за големи данни*. Терминът **големи данни** (Big Data) е неизбежен в света на технологиите днес. Разнообразието от инструменти и технологии в областта на големите данни нараства изключително бързо. Анализът и работата с големи данни изискват специфични знания и умения.

Авторките Катя Рашева-Йорданова и Стефка Толева-Стоименова ни потапят дълбоко в света на големите данни с теоретико-концептуално изследване на литературните източници, моделирането на компетентностния профил на експертите в областта на науката за данните чрез изследване на пазарните нужди в България и създаването на учебен план за магистърска програма, която ще изгражда такива специалисти.

Монографията може да послужи за въведение в света на големите данни и науката, която ги изучава. Авторките са се стремили да представят пълен набор от дефиниции на термина **големи данни**, както и възникналите проблеми и предизвикателства в областта, свързани с жизнения цикъл на данните, процесите и управлението им. Те се спират подробно на всеки етап от обработката на данни – методите, моделите и технологиите, които може да се използват за управлението им.

В монографията са поставени основите и на представянето на **науката за данните** и нейните цели, задачи, принципи и методи. Обърнато е специално внимание на комплексността и мултидисциплинарността на тази наука. Представените различни схващания за дисциплините, влияещи на новоизграждащата се наука, осигуряват широк поглед на читателя върху многообразието в тази област.

Авторките са млади учени от екипа на Факултета по информационни науки на Университета по библиотекознание и информационни технологии. Екипната им работа е пример за синергията, която колективите, които работят с големи данни, трябва да притежават. Както те самите споделят:

С оглед на комплексния характер на науката за данни специалистите по данни трябва да притежават комбинация от умения и компетенции в областта на програмирането, анализа, статистиката, моделирането и комуникацията. Те използват различни източници (...) с цел извличането на полезна информация с добро качество, точност, достоверност и др. Затова трябва да притежават обща компетентност за работа с данни и да отговорят на следните въпроси (...) (с. 79 – 80).

Катя Рашева-Йорданова е експерт по информационни и комуникационни технологии, а Стефка Толева-Стоименова – по статистика и математическо моделиране – математика като цяло.

Един от основните приноси в монографията е проучването на нуждите на пазара от специалисти по големи данни, и то с фокус върху това какво е търсенето в нашата страна и какви са нуждите на българския пазар.

Работодателите правят разлика между младши учен по данни (Junior), специалист по данни от средно ниво (Mid-Level) и старши специалист по данни (Senior).

Основното различие между трите равнища в кариерното развитие е свързано, от една страна, с възможностите за самостоятелност и независимост при изпълнение на поставени задачи. От друга страна, различието е провокирано от възможностите за организация на работния процес в съответствие с нуждите и сроковете, поставени от компанията.

Анализът на проучването, който авторките са направили, е съществен за следващия етап – създаването на профила на специалиста по данни, а оттам и изграждането на магистърска програма.

Рашева и Толева са *обходили* Мрежата и са се запознали със съществуващи магистърски и бакалавърски програми по света. Анализирали са задълбочено какво се изучава в тях, и са използвали подхода на компетентностно базираното обучение при създаването на ме-

тодика за проектиране на свой вариант на програмата **Анализ и управление на данни**.

Методиката насърчава непрекъснато усъвършенстване чрез разглеждане на стратегическите планове на университета. Тя се основава на интензивното изследване и анализ на потребностите на развитието и обучението, основано на компетенции (...)

Моделът на професионални компетенции за анализ на големи данни трябва да служи като основна отправна рамка при всички изследователски процедури, включително:

- *поставянето на целта и оценяването на компетенциите;*
- *проектирането на учебното съдържание и моделирането на взаимодействията в процеса на обучение;*
- *процедурите на контрол.*

В монографията е поставено добро начало и за създаване на речник в областта на големите данни за българската аудитория. Той може да бъде разширен, разбира се, но е основа за бъдещите специалисти в областта.

Както самите автори акцентират:

Създаването на специалисти е труден и продължителен процес, в който ролята на университетите е свързана с предлагането на модерно образование и формирането на специалисти за конкретни сектори и отрасли на икономиката.

Монографията е само началото на техния път в големите данни и изграждането на експерти в областта. Авторките са извървели дълъг път, поставили са си амбициозни задачи, на които са отговорили, като са вложили своите знания и опит и са изградили синергия, каквато е нужна на специалистите по науката за данните.

Доц. д-р Евгения Ковачева

Катя Христова Рашева-Йорданова
Стефка Янкова Толева-Стоименова

**ИЗГРАЖДАНЕ НА DATA SCIENCE КОМПЕТЕНТНОСТ
ЗА ПРЕОДОЛЯВАНЕ
НА ДИГИТАЛНОТО НЕРАВЕНСТВО**

Монография

Българска
Първо издание

Научни рецензенти:
доц. д-р Волин Карагъзов
доц. д-р Евгения Ковачева

Редактор и коректор
Жана Ганчева

Превод на английски език
д-р Веселин Чантов

Графичен дизайн на корицата
Ралица Йотова

Формат 60/90/16
Печатни коли 17,5
Предпечат и печат БПС ООД

Академично издателство „За буквите – О писменехъ“
ISBN 978-619-185-392-2

София, 2019