

Abstracts of Dissertations

Institute of Information and
Communication Technologies

BULGARIAN ACADEMY OF
SCIENCES



9 / 2023



Application of data
science
in the virtual
educational space

Daniela Orozova

Приложение на
науката за данните
във виртуалното
образователно
пространство

Даниела Орозова

Автореферати на дисертации

Институт по информационни и
комуникационни технологии

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ISSN: 1314-6351

Поредицата „Авториферати на дисертации на Института по информационни и комуникационни технологии при Българската академия на науките“ представя в електронен формат авториферати на дисертации за получаване на научната степен „Доктор на науките“ или на образователната и научната степен „Доктор“, защитени в Института по информационни и комуникационни технологии при Българската академия на науките. Представените трудове отразяват нови научни и научно-приложни приноси в редица области на информационните и комуникационните технологии като Компютърни мрежи и архитектури, Паралелни алгоритми, Научни пресмятания, Лингвистично моделиране, Математически методи за обработка на сензорна информация, Информационни технологии в сигурността, Технологии за управление и обработка на знания, Грид-технологии и приложения, Оптимизация и вземане на решения, Обработка на сигнали и разпознаване на образи, Интелигентни системи, Информационни процеси и системи, Вградени интелигентни технологии, Йерархични системи, Комуникационни системи и услуги и др.

Редактори

Геннадий Агре

Институт по информационни и комуникационни технологии, Българска академия на науките
E-mail: agre@iinf.bas.bg

Райна Георгиева

Институт по информационни и комуникационни технологии, Българска академия на науките
E-mail: rayna@parallel.bas.bg

Даниела Борисова

Институт по информационни и комуникационни технологии, Българска академия на науките
E-mail: dborissova@iit.bas.bg

Настоящото издание е обект на авторско право. Всички права са запазени при превод, разпечатване, използване на илюстрации, цитирания, разпространение, възпроизвеждане на микрофилми или по други начини, както и съхранение в бази от данни на всички или част от материалите в настоящето издание. Копирането на изданието или на част от съдържанието му е разрешено само със съгласието на авторите и/или редакторите

The series Abstracts of Dissertations of the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences presents in an electronic format the abstracts of Doctor of Sciences and PhD dissertations defended in the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences. The studies provide new original results in such areas of Information and Communication Technologies as Computer Networks and Architectures, Parallel Algorithms, Scientific Computations, Linguistic Modelling, Mathematical Methods for Sensor Data Processing, Information Technologies for Security, Technologies for Knowledge management and processing, Grid Technologies and Applications, Optimization and Decision Making, Signal Processing and Pattern Recognition, Information Processing and Systems, Intelligent Systems, Embedded Intelligent Technologies, Hierarchical Systems, Communication Systems and Services, etc.

Editors

Gennady Agre

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
E-mail: agre@iinf.bas.bg

Rayna Georgieva

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
E-mail: rayna@parallel.bas.bg

Daniela Borissova

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
E-mail: dborissova@iit.bas.bg

This work is subjected to copyright. All rights are reserved, whether the whole or part of the materials is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. Duplication of this work or part thereof is only permitted under the provisions of the authors and/or editor.



BULGARIAN ACADEMY OF SCIENCES

Abstract of Doctor of Science Thesis

Application of data science in the virtual educational space

Daniela Ananieva Orozova

Approved by Scientific Jury:

Acad. Ivan Popchev
Acad. Krassimir Atanassov
Cor. member Lyubka Doukovska
Prof. Veselin Videv
Prof. Radoslav Yoshinov
Prof. Maria Nisheva-Pavlova
Prof. Maria Hristova



**INSTITUTE OF INFORMATION AND
COMMUNICATION TECHNOLOGIES**

Department of Intelligent Systems

INTRODUCTION

1. Data science in Learning

Virtual educational spaces [3] are informational and social spaces that integrate heterogeneous technologies and different pedagogical approaches. They are a medium for delivering learning materials and educational services to different target groups, regardless of time and space. The data accumulated by the work of the educational space is constantly increasing. Many countries and universities around the world are building infrastructures to analyze this data. Siemens & Long [90] define *Data analytics in Education* as "the measurement, collection, analysis and reporting of data about learners and their contexts in order to understand and optimize learning and the environment in which it occurs." Chatti [23] defines the purposes of this analysis such as: monitoring, prediction, individualization, intervention in the learning process, evaluation and recommendations of the learner.

The terms *Big Data*, *Data Analytics* and *Data Mining* describe both the data itself and the technologies for data collection, processing, management and analysis methods [24]. *Data Mining* is the process of searching for hidden data and regularities, previously unknown, non-trivial and practically useful, necessary for making decisions in various spheres of human activities. *Big Data Analytics* is a development of the *Data Mining concept*. It is also a development of solved tasks, fields of application, data sources, processing methods and technologies. In 1998, Hayashi Chikio [44] introduced the term *Data science* as a new, interdisciplinary concept, with three aspects: data design, collection and analysis. *Data science* combines multiple approaches and techniques related to data analysis, knowledge discovery, machine learning, artificial intelligence, programming, communication, etc. *Data science* is an "alloy" of various disciplines, technologies and tools for data analysis and is the main means of discovering and exploiting the potential of Big Data [26].

The term "*Data analysis*" refers to the processing of data through conventional (classical statistical, empirical or logical) theories, technologies and tools to extract useful information and for practical purposes. The term "*Data analytics*", on the other hand, refers to the theories, technologies, tools and processes that enable a deep understanding of the studied data [21].

Data collection for learning analysis refers to the entire process and includes all data obtained during learning activities. These are many and varied types of data. The international standardization organization, IMG Global Learning Consortium (IMG Global) classifies the data that can be collected and analyzed in the field of education into five types: learning content data; data on the educational activity; operational data; data related to career development; learner profile data. This is a prerequisite for the emergence of *Learning Analytics* [23], a new branch of data analysis that includes goals and methods drawn from educational and psychological research.

In 2009, Avinash Kaushik published the book "Web Analytics 2.0", in which he defined *Digital Analytics* as "the continuous analysis of quantitative and qualitative data from the web space to improve the online experience of potential customers and thus lead to the desired results". [51] An approach of dividing the data into different principles has been introduced, which gives them a specific meaning. The entry and popularization of social networks marks the beginning of modern social analysis *Social Analytics*, defined as "observation, analysis, measurement and interpretation of digital interactions and relationships of people, topics, ideas and content" [92]. This analysis examines the role of social interactions in the learning process and their impact on learner progress.

Educational Data Mining is defined as "the discipline concerned with developing methods for examining the unique and increasingly large-scale data that come from educational environments and using those methods to better understand of learners and the environment in which they learn" [47]. Integrating the means of extracting knowledge from data in educational environments, trends in the development of e-learning processes and its service can be sought. Following the process of selection and use of different means, optimal learning environments can be built with opportunities for personalized

mastery of key knowledge, skills and competencies. They can predict the potential loss of learners, determine which activities will be preferred and how effective they will be. New information technologies contribute to research, analysis and development of learning environments by monitoring and measuring various aspects of the virtual and physical environment in which learning takes place.

2. Purpose and tasks of the dissertation work

The purpose of the research is theoretical summaries of the processes of observation and analysis of data from the dynamic interaction of objects and the creation of methods and models for solving scientific or applied scientific problems in the virtual educational space.

To achieve the goal, the following four **tasks are set**:

- Analysis of the application of tools for extracting knowledge from the data in the learning spaces and search for solutions for the personalization of e-Learning and distance learning.
- Creation of methods for evaluating and predicting the knowledge, skills and competences of learners in the virtual educational space.
- Creation of models, as a result of theoretical summaries of the processes of observation and analysis of the activities of learners related to Big Data Analytics, Data Mining, Web metrics, Generalized net, Machine Learning and Fuzzy logic.
- Introduction of basic modules and tools of Data Science for solutions of applied scientific problems in education.

The dissertation on "Application of data science in the virtual educational space" includes 49 figures, the bibliography covers 220 sources, of which 197 are in English.

The text is organized into an introduction, five chapters and a conclusion. The **introduction** provides an overview of the main concepts of the researched area. The purpose of the dissertation work and the tasks related to its achievement are defined. The **first chapter** examines the field of virtual educational space, the main characteristics and possibilities of this type of infrastructure and the motivation for the conducted research work. **The second chapter** presents the status, proposed solutions, general characteristics and opportunities for the integration of means and tools to the virtual educational space. In **the third chapter**, methods and techniques are proposed for modeling the processes in the educational space, with the application of data mining tools in an educational context (Educational Data Mining). The **fourth chapter** examines work processes in the virtual educational space and their formal models via Generalized nets. **The fifth chapter** presents practices and trends in the entry of Data Science into higher education. **The conclusion** summarizes the results and gives directions for further research on the topic of the dissertation work. Attached are: Declaration of originality of results, Contributions of the dissertation, List of publications on the topic of the dissertation.

CHAPTER 1. VIRTUAL EDUCATIONAL SPACE

1.1. Data architecture of a virtual educational space

Examining the architecture of a virtual educational space from a data perspective, emphasis is placed on the various levels of data representation and associated mechanisms for data collection and analysis, for information extraction and prediction of environment and learner behavior. Figure 1 summarizes four basic levels of data representation, but other levels may be considered depending on the storage and access criteria applied.

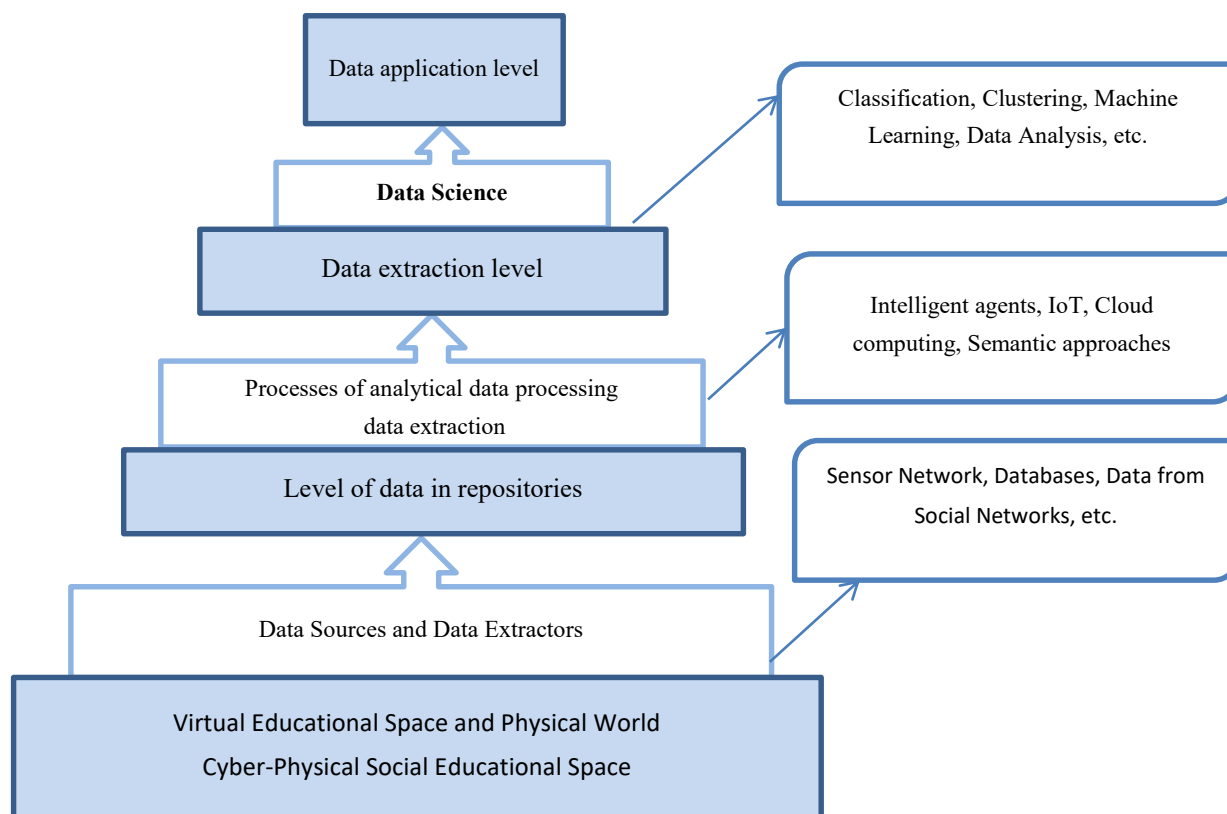


Figure 1. Functional levels in the virtual educational space

First level. *Data from the virtual educational space and the physical world.* The sources cover a wide field, combining data obtained from sensor sensors, data from social networks, a database of the field of learning and many others.

Second level. *Data repositories.* This level provides data about the state of the physical world transferred to the virtual environment of space. It includes the use of technologies for data collection, storage, processing and analysis tools.

Third level. *Repositories with analytic subspaces of the data.* Data in virtual learning spaces can be viewed from different perspectives and subjected to different types of analysis. **Fourth level.** *Building applications.* The application layer provides users with a set of tools and interfaces for managing, processing and visualizing data from previous layers.

New technologies for working with big data and the technical power of modern computers create an opportunity for innovative applications based on the integration of heterogeneous data.

1. 2. Virtual educational space and the concept the concept Internet of things

The concept of the Internet of Things defines the devices or objects (things) constituting an ecosystem that they inhabit and enrich by sharing their functionalities, content and knowledge. E-learning must be included in the IoT environment to keep up with user demands and the evolution of the Internet. Every aspect of online education is like something or a combination of things in this ecosystem - learning, assessment, intelligent educational assistance. The ubiquitous presence of the Internet in our lives and its gradual transformation into the Internet of Things is a prerequisite for the creation of "cyber-space" [16]. In [14], a "cyber-physical system" (Cyber-Physical-System, CPS) is defined as an engineering system made up of computational and physical components. "Physical" are the elements of the system occupying physical space, "cyber" are the computing and communication elements of the system. Accounting for the social component in CPS transforms it into a Cyber-Physical Social System (CPSS) [102]. The implementation in the field of education appears as CPSES (Cyber-Physical Social Educational System) and represents a subset of CPSS in the "education and training" domain.

Successful results for integration of the virtual environment with the physical world can be indicated. In the laboratory "DeLC (Distributed eLearning Center)" of the University of Plovdiv, се разработва кибер-физическа-социална система is being developed, which is being built as a successor to the DeLC e-learning environment [93], It is called Virtual Educational Space (VES) and is implemented through an agent-oriented approach. The constructed educational space [95] is "populated" only by active autonomous components called assistants. Assistants, implemented as intelligent software agents, constitute the core of the space, which support the planning, organization and execution of the learning process [95]. Space integrates the virtual world of e-learning with smart devices, sensors to collect data from the physical world, communication infrastructures and system architecture.

Most universities in the country implement projects for the application of information technologies and innovative approaches based on ready-made learning management systems. These environments lack close integration of the virtual environment with the physical world of the learning process.

In the framework of this dissertation, the virtual educational space is defined as an environment that integrates different information technologies and pedagogical approaches for the delivery of learning materials and educational services in an abstract university (without realizing integration with the physical world). It covers various educational tools and resources that are used in digital form to enhance the educational process. Efforts are in the direction of developing a methodology for structuring and modeling the learning processes, detailing and formalizing the processes.

1. 3. The virtual educational space and Big Data

The virtual education space collects a huge amount of data daily, which requires a new approach to data processing. The term "BigData" concentrates efforts in the organization, storage, processing and analysis of huge masses of data that are so voluminous and complex that it is impossible to process them with traditional data processing applications [12] . In [62], a model is proposed describing the processes related to storage, processing and analysis of Big Data, which require a new look and joint application of a number of established technologies. In [63], a model of the processes related to distributed computing is proposed, based on the Map/Reduce paradigm, using different algorithms and tools for their implementation.

1.4. The virtual educational space of emerging technologies and risks

Leading technologies that are characterized by radical novelty, rapid growth and impact on other technologies are called *emerging technologies* or *disruptive technologies* [19]. In American sources, these technologies are also defined as *brilliant technologies* [20] that are used and developed over time and have the potential to increase their impact on knowledge production processes. **Artificial intelligence is the "heart" of emerging technologies**, because the scientific breakthroughs related to it form directions, the functioning of which depends to the greatest extent on the presentation of knowledge and imitating the abilities of human reasoning.

Digital technologies are becoming increasingly complex and integrated, and as such they are causing significant transformations in society and the economy. This change has *positive impacts*, but we must also consider *negative impacts* such as: loss of jobs in traditional professions; cybercrime and hacker attacks; algorithmic errors and incomprehensibility; need for new concepts of responsibility, accountability and governance; increasing inequality between people. Those exposed to the greatest number of risks are the technologies of "Artificial intelligence and decision making", "Big data and solutions", "Sharing economy", "3D printing and manufacturing, consumer products and healthcare", "Our digital presence" and "Autonomous cars" [84].

To continue to develop the concept of the virtual educational space in this dynamic and heterogeneous environment, it must adapt to the characteristics and requirements that the environment imposes. But any changes must be subordinated to policy and investment for reliable artificial intelligence [85] and based on an ethical and human-centered approach [31].

1.5. Models for analysis of the subject area

In the literature [4], formal software analysis models are divided into: property-based models, executable models, and models integrating both approaches.

Among executable models, Petri nets are considered as a simple and clear modeling tool [76]. They have many extensions adding a number of new properties and modeling capabilities. Regardless of their specifications, all of these tools have positions and transitions (indicated by arcs) and tokens that move on the grid. Main extensions of Petri nets are [8, 9] :

- o EN – *Evaluation nets* add token movement duration;
- o TPN – *Temporal PNs* add transition activation moment;
- o CPN – *Color PN*. In colored Petri nets, each token has a color and can only move along arcs of the same color;
- o SPN – *Stochastic PN*. Arc selection in stochastic Petri nets is based on a randomly generated number;
- o SMPN – *Self Modifying PN*. In self-modifying Petri nets, arc selection is based on a generated binary number 0 or 1, (with 0 being no transition);
- o PRON – *Pro-Net* introduces transition type;
- o PTN – *Predicate/Transition Net* . These networks define a transition condition;
- o MN - *M Net*, where the tokens also perform additional procedures;
- o GMPN – *Generalized Modifying PN* – the kernel is absorbed when the predicate generates a 0, otherwise when a 1 is generated the kernel performs the transition.
- o GN *Generalized Net* – unite the different extensions into a single formalism for describing parallel processes. The different extensions of Petri nets and the relationship between them are presented in Figure 2.

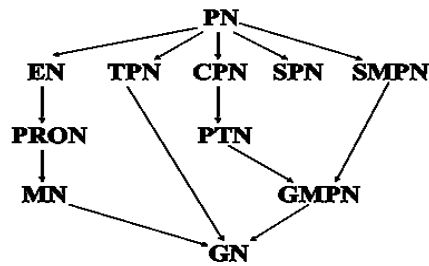


Figure 2. Extensions of Petri nets

The dissertation follows a brief description of the concept of Generalized Nets (GNs), presented in [8, 9], a formal description of transition and a formal description of GMNs.

In the fourth chapter of the dissertation, formal models of processes and components of the virtual educational space are presented through the apparatus of generalized nets. They represent a summary of a created series of models related to monitoring and analyzing the activities of learners in an educational environment, presented in the author's publications [59, 62, 63, 68, 40, 41, 70, 71, 72, 74, 69, 97, 99].

CHAPTER 2. MEANS OF THE VIRTUAL EDUCATIONAL SPACE

Virtual educational space is considered as an environment that integrates various information technologies, pedagogical approaches, educational tools and resources that are used in digital form, to deliver learning materials and educational services in an abstract university.

Data science can provide multiple benefits in university management and help improve efficiency and decision-making. Considering the specific needs and goals of the educational institution, appropriate tools and methodologies for data analysis can be selected. Space is viewed as an active system open to components with a common approach to accessing and using information for analysis and decision-making [1, 55]. Such a system provides basic components such as:

- Maintains means of collection, purification and efficient storage of the data. Stored information is accessible at different levels of access and by different applications for retrieval, display and decision making, as well as data access control. Interrelates information obtained from various sources.
- Provides knowledge management tools with intelligent search and reading capabilities, data integrity and correctness control, automated information extraction, etc. This includes data quality management tools and automated data pattern discovery and self-updating data capabilities.
- Optimizes the process of fulfilling requests when using the information. Transforms and visualizes information to be easily perceived and used by the user;
- Възможности за интегриране на компоненти с разнородни технологии и педагогически методи, посредством възприемане на общи архитектурни подходи за системата.

In addition to basic resources, a way to engage learners' attention is to integrate into the learning process those resources that they use daily in their personal virtual space.

Integrating knowledge mining tools into learning environments can significantly improve the educational process by providing valuable information to learners, educators, and administrators, but in turn requires effective management of educational data. This includes storing, processing and protecting the data, ensuring its security and confidentiality. They must be used in accordance with ethical principles and respect the rights of the individual and privacy.

This chapter of the dissertation summarizes the results of the conducted research and analysis of the main characteristics of a virtual educational space on University, presented in detail in the publications [61, 79, 82, 94, 96, 97]. Advantages of standardized e-learning content in the delivery of educational services are substantiated. The need to apply data mining tools in learning spaces and search for solutions to personalize e- and distance learning has increased.

CHAPTER 3. MODELS FOR DATA ANALYSIS IN THE VIRTUAL EDUCATIONAL SPACE

This chapter of the dissertation summarizes the results of the conducted experiments and analyzes of the collected data for learners in various forms of e-learning, presented in detail in the author's publications [75, 79, 80, 82]. In the conducted research, a link between statistical methods, machine learning, behavioral pattern discovery and data analysis is sought. The developed algorithms and software are presented in [57, 58, 83]. Models are proposed as a result of theoretical summaries of the processes of observation and analysis of learner activities related to Big Data Analytics, Data Mining, Web metrics, Machine Learning and Fuzzy logic.

Data analytics in a virtual educational space

Various forms and methods of testing are used to assess the students' knowledge, such as: online or offline tests, answers to open questions, testing and reading of code, solving tasks, etc. The final assessment in the discipline is often complex and includes several components involved in its formation with different weights. It must reflect the various aspects of the student's training (theoretical knowledge and practical skills), be tailored to the specifics of the subject, age and individual characteristics, etc.

An experiment on evaluation and analytical studies of learner data is presented in [75]. For the purposes of the analysis, we consider a specific approach to student assessment, where the grade at the end of the semester is defined as a weighted average \bar{x} of all grades during the semester. Each rating has an importance or weight. The function used in the particular case has the form:

$$(1) \quad \bar{x} = \frac{\sum xw}{\sum w}$$

where x is the score, w is the weight of each score.

The assessment of students in the Mathematics subject at the end of the semester is determined based on the following tests: *Test 1 Sets*, *Test 2 Logic*, *Test 3 Geometry*, *Test 4 Statistics*, *Test 5 Probability*, *Final Exam*, *Quizzes*, *Homework*, *Projects* and *Class Activities*.

Each criterion has a degree of importance and the function has the form:

$$(2) \quad \bar{x} = \frac{\sum T1*0.09+T2*0.09+T3*0.09+T4*0.09+T5*0.09+FinEx*0.15+Act*0.40}{\sum w}$$

where \bar{x} is the grade at the end of the semester, w is the weight of each grade. We look at how individual test scores affect grading and how in-class and extracurricular activities affect the end-of-semester grade. In addition to the points accumulated for the individual components, we also store data for each student: Age (years is cold), Gender (gender), New Student Experience (NSE), FullTime-PartTime (credits for the subject are 12 or more), Student Program (type of study), Times Taken Course (for which time the student studies the subject).

3.1. Using Spreadsheets for analytical studies of learner data

Spreadsheets (especially Excel) offer effective functions for summarizing data. Filters *provide* options for selecting records. Conditional formatting *colors* data that meets certain criteria and helps us detect deviations and trends in the data. Charts visually *show* deviations and trends [30]. Using Excel tools, we performed an initial analysis of student mathematics performance outcomes over five semesters. The graph of Fig. 3 shows a good absorption of the learning material. The lowest scores were on Test 4 Statistics, where 75% of learners scored lower than 50% of learners on Test 1 Sets, Test 2 Logic, Test 3 Geometry and Test 5 Probability.

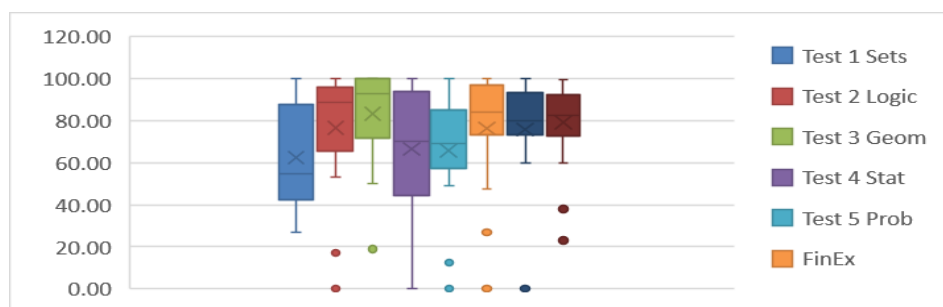


Figure 3. Graph regarding the distribution of grades by modules

The change in teaching strategies had an impact on how learners absorbed the material. As a result of the research, a recommendation was made to change the sequence in the study of two disciplines.

3.2. Data analytics in the learning space with the Orange system

Various data mining software tools are available today. Among the most popular are: RapidMiner, RapidAnalytics, WEKA, PSPP, KNIME, Orange, Apache Mahout, jHepWork, Rattle, GhostMiner, XENO, SAS Enterprise Miner, Polyanalyst and IBM SPSS modeler. For the purposes of this study, a tool was developed to identify and predict the reasons for students falling behind or dropping out, presented in [78]. The “*Orange Data Mining system*” [56], an open source machine learning and data mining software written in Python, was used.

Supervised learning

The main task of supervised learning is to create a model of labeled data, which allows making predictions about future data. The main techniques are: classification when the class labels are discrete and regression when the score is a continuous value. A number of tools for building classification and regression models have been built into the *Orange Data Mining System*. In the next subsection, experiments are presented to evaluate and predict the knowledge of learners in an educational space using the tools: Logistic Regression, Naïve Bayes Classifier, Support Vector Machines (SVM), Decision Trees and Artificial Neural Networks, etc.

3.2.1. Problems in preparing data for analysis and approaches to solving them

For data analysis in an educational space, the main input data are the points obtained by the learners on different assessment components, and the output is the corresponding final grades set by the teacher. Main tasks for data preparation for analysis are [22]: data cleaning (Data Cleaning), data integration (Data Integration), data transformation (Data Transformation), data reduction (Data Reduction). Data cleaning is the process of ensuring that data is correct, consistent and usable. *Duplicate* data can occur when combining datasets from multiple sources. *Irrelevant* data is data that does not fit the problem being solved. *Structural errors* in the data may occur during measurement or data transfer. *Outliers* can cause problems with some types of models. When we do not have the complete set of data for a given characteristic, we have missing *data*. The process of preparing data includes identifying errors, correcting, deleting, arranging or otherwise processing, keeping only the potentially useful data.

3.2.2 Evaluating and predicting the knowledge of learners in the virtual educational space

It is proposed here as a **multi-step process**, a method for evaluating and predicting the knowledge, skills and competences of learners in the virtual educational space [66]. **The method includes five steps** and can be adapted to different *Emerging technologies courses* in the virtual educational space.

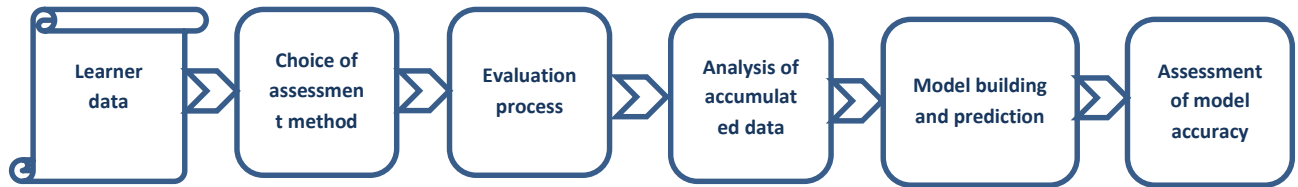


Figure 4. Multi-step process of analyzing and predicting learner knowledge

Step 1. Choice of assessment method. Determination of key knowledge and competences for the studied technology. Determining the degree of severity (importance) of each analyzed competency.

- Basic **theoretical knowledge** is assessed using components such as: *intermediate tests, problem solving and case studies, exams, summary discussion* etc. These components assess the acquired knowledge and the ability to understand the theoretical material studied. Open test questions, multiple choice questions, listing and comparing objects, giving examples of concepts, explaining and using algorithms are used. As well as questions related to explaining, interpreting and visualizing solutions.

- The main **competencies** are assessed by the students' ability to apply the acquired knowledge to make non-standard decisions in: *control and homework, course assignments and projects*, in which new tasks are solved, critical analysis of decisions is carried out, potential risks are determined, independent conclusions and conclusions are made.

Step 2. Evaluation process. Assessment procedures are also conducted during the overall learning process. The results are accumulated and stored for the purpose of student profiling, subsequent analysis and prediction in the assessment of new students.

The obtained results look for dependencies in the individual assessment components between theoretical and practical knowledge, skills and competences. In some cases, a student gets a high grade on the tests and a low grade on the test, or vice versa. Creating an automated scoring algorithm when specific values of the scoring components are available requires a non-standard solution.

Step 3. Analysis of the accumulated data from the conducted trainings in a real environment. Various machine learning algorithms are available. The main input data are the points obtained from the evaluation components, and the output is the corresponding final grades of the students in the discipline. In training methods, part of the sample data is used for training the algorithm and another part is used for testing. If the test results are not good, the training process can be repeated or it can be judged that the chosen approach is not good for solving the specific problem.

Step 4. Building a prediction model. Based on the accumulated data, a classifier is created that makes predictions for current grades in the discipline. Data analysis systems such as: SPSS, Orange, Weka and others offer tools based on algorithms for: decision tree, logistic regression, Bayes theorem, neural networks and a number of others. After training is complete, the implemented models can be applied to new data.

Step 5. Evaluation of the prediction accuracy of the algorithms. After training, testing is performed for accuracy and precision of the model's performance.

How good a classifier is is determined by the values of relevant quality assessment metrics. It is not enough to watch only one of them. Several metrics should be considered, and the choice of which are more important depends on the tasks and the objectives involved.

Confusion Matrix is one of the most popular ways to evaluate the quality of the classification [89]. Represents an $N \times N$ matrix, where N is the number of classes of the target variable. By applying this tool, the classifier determined and the actual values can be compared according to 4 metrics: True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN).

- *Overall accuracy* is a metric that gives information about what proportion of all cases are correctly classified:

$$(3) \quad Accuracy (ACC) = (TP+TN) / (TP+FP+FN+TN)$$

The total number of correctly classified objects is divided by all cases.

- *The total error* shows what fraction of all objects the classifier assigned to the wrong classes:

$$(4) \quad ERR = (FP+FN) / (TP+FP+FN+TN) = 1 - ACC$$

It can be calculated by dividing the total number of misclassified objects by all cases or by subtracting the *Accuracy score* from 1.

It is necessary to take into account the metrics: *Precision*, *Recall*, *Specificity* and *F1-score*. These metrics are the most popular and commonly used when evaluating a classifier:

- *Precision* - the metric shows what fraction of objects classified as positive are actually positive:

$$(5) \quad Precision = TP/(TP+FP)$$

- *Recall* or True Positive Rate (TPR) - gives information about how much of the positive class was detected by the classifier:

$$(6) \quad Recall = TP/(TP+FN)$$

- *Specificity* or True Negative Rate (TNR) - shows how much of the negative class was found by the classifier:

$$(7) \quad Specificity = TN/(TN+FP)$$

- *F1-score* - summarizes *Precision* and *Recall* into a single value. The value of this indicator is maximum when *Precision* and *Recall* are equal:

$$(8) \quad F1-score = 2*(Recall*Precision)/(Recall+Precision)$$

The analysis of the results is related to decisions by the managers of the training process, how to stimulate the learners and how to help the learners at risk: additional exercises and tasks, additional individual and team work, step-by-step explanations, work with software such as modeling tool etc.

Application of the evaluation and forecasting method was carried out in the discipline "**Artificial Intelligence**" using the tools of *the Orange system* for experimentation and conclusions [66]. The presented discipline (with Professors Academician Ivan Popchev and Prof. Dr. Daniela Orozova) is part of the virtual educational space. It is structured in four modules: Artificial Intelligence – characteristics and issues; Search for a solution in the state space; Presentation of knowledge; Smart decision making. In the e-learning environment for the discipline, there are provided materials with a volume of 587 MB of textual description, examples and 14 links with useful links. Moodle and Microsoft Teams environments are used in teaching. This discipline is the basis of the courses: "Analysis and design of databases and knowledge", "Knowledge management in computer systems", etc.

Step 1. For the evaluation of the students in the discipline "Artificial Intelligence", evaluation components have been defined, each of which inspects theoretical knowledge, practical skills and competences with a different cognitive level. 3 tests are defined (3 x 5 = 15 points), quizz Test (15 points), Project - 45 points and Summary Discussion - 25 points.

In *the project*, each student chooses a topic from the main areas of artificial intelligence, such as: Ontologies engineering, Semantic Web, Knowledge representation, Computational intelligence, Robotics, Natural language processing, Machine Learning, Deep learning, Soft computing, Pattern recognition,

Multi-agent systems, Artificial neural networks, Genetic algorithms, Knowledge based systems, Decision support systems, Business intelligence, Data Science, Fuzzy sets and systems, E-learning, etc. The project necessarily includes: status, development trends, identification, analysis and evaluation of the selected toolkit for impacting potential risks, unsolved problems, conclusions, conclusion and bibliography.

The general discussion with the student is on the subject of the project, risk management, monitoring, control and assessment of risk management of potential risks and the possibilities of solving new tasks with non-standard solutions.

The example model used for assessment in the discipline is presented in Fig.5. This assessment model can be dynamically changed and adapted to the specific discipline. For example, in the discipline "**Knowledge Management in Computer Systems**" an alternative model is applied .

Evaluation component	Meaning	Rating scale
Theoretical knowledge module 2	Test_1	up to 5 points
Theoretical knowledge module 3	Test_2	up to 5 points
Theoretical knowledge module 4	Test_3	up to 5 points
Practical competences	Test	up to 15 points
Practical knowledge and skills	Project	up to 45 points
Theoretical knowledge and skills	Summary discussion	up to 25 points
Final assessment	Final assessment	Score points [2, 6]

Figure 5. Example model for assessment by discipline

Step 2. Conduct training and evaluation process. The task is related to finding a general approach for automated assessment and prediction of student grades. To reduce subjectivity in the assessment of practical skills, the assessment is allowed to be carried out by an external assessor from companies in the field of information technology such as: Technologika, Scale Focus, etc. For each discipline, students can receive up to 100 points and the final grade is formed according to the following scale: from 54 to 60 points - *Average* (3); from 61 to 70 points - *Good* (4); from 71 to 80 points - *Very good* (5); from 81 to 100 points - *Excellent* (6).

Step 3. Analysis of the accumulated data from the conducted trainings in a real environment.

In the process of work, many experiments were conducted. The main goal is to solve a classification problem by determining whether it is possible to predict the score (output variable) using the input variables (the points of the individual score components) that are stored in the model. Various techniques are applied to solve the classification problem, using the tools of *the Orange Data Mining System*.

At the beginning, a workflow is created and through the "*File*" tool **we load the data** for the students' evaluations on the various evaluation components in the number of points. They can be entered from Excel (.xlsx), a tabbed text file (.txt), a comma-separated data file (.csv), or a URL. For better understanding, the data can be visualized by some columns or extracts from them. For example, one can link the data file to the *Scatter Plot tool* and select the columns whose values will be plotted on the X and Y axes, the colors used, shapes, sizes and other parameters. Another popular data visualization tool is "*Distribution*", which can be used to show a distribution in the data set by a given attribute.

Depending on the purpose and type of data, we choose a specific regression or classification tool and **set the target variable**. Next is **data cleaning**. Different approaches can be taken: delete the missing values or replace them appropriately. Using the "*Impute*" tool (Fig. 6), one chooses among different imputation methods. The default is the "*Remove the rows with missing values*" option. Other possible options are: *Distinct Value, Random Values, Model-Based*.

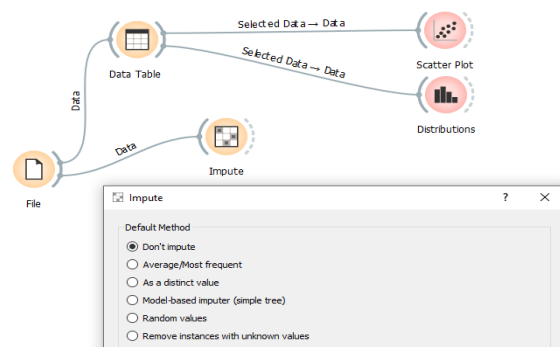


Figure 6. Applying data loading and cleaning tools

Step 4. Building and training prediction models. In the experiment, the tools of the system are consequently applied: **Tree**, **Random Forest**, **Logistic Regression**, **Naïve Bayes**, **Support Vector Machines (SVM)** and **Neural Network**. The workflow for creating and training the models is shown in Fig. 7.

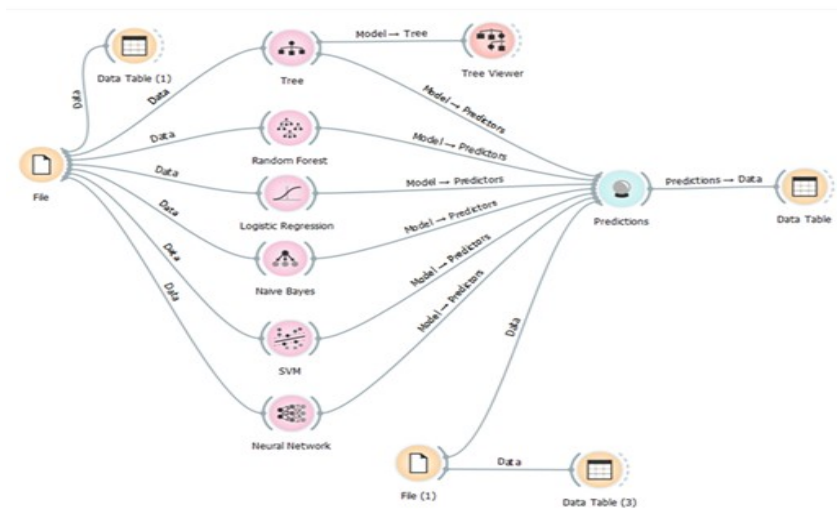


Figure 7. Model training and prediction workflow

Anticipating new data. Through the trained algorithms, when setting a new combination of values for the selected components, the output variable is expected to be determined. At this step, we consider the model ready for practical application. The model gains independence and draws its own conclusions based on data sets and training. Figure 7 shows the prediction workflow using the *Prediction tool* of the Orange system. The new data is fed through a *Test.xlsx file* that has the same structure as the original data table, but the output variable column is not set. A general view of the obtained prediction result is shown in Figure 8. These are the prediction results for the final grade of learners obtained by the different models.

Data & Predictions					
	Tree	Logistic Regression	Random Forest	SVM	Naive Bayes
1	0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Excellent	0.00 : 0.49 : 0.02 : 0.06 : 0.43 → Excellent	0.00 : 0.57 : 0.11 : 0.00 : 0.32 → Excellent	0.01 : 0.18 : 0.03 : 0.01 : 0.77 → Very Good	0.00 : 0.85 : 0.01 :
2	0.25 : 0.00 : 0.50 : 0.25 : 0.00 → Good	0.01 : 0.00 : 0.45 : 0.22 : 0.32 → Good	0.20 : 0.10 : 0.20 : 0.50 : 0.00 → Middle	0.22 : 0.02 : 0.23 : 0.51 : 0.03 → Middle	0.39 : 0.00 : 0.04 :
3	0.25 : 0.00 : 0.50 : 0.25 : 0.00 → Good	0.00 : 0.00 : 0.20 : 0.30 : 0.50 → Very Good	0.17 : 0.00 : 0.76 : 0.05 : 0.02 → Good	0.03 : 0.05 : 0.42 : 0.08 : 0.42 → Very Good	0.57 : 0.00 : 0.22 :
4	0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Excellent	0.00 : 0.72 : 0.03 : 0.04 : 0.22 → Excellent	0.00 : 0.60 : 0.20 : 0.00 : 0.19 → Excellent	0.01 : 0.43 : 0.03 : 0.01 : 0.52 → Very Good	0.00 : 0.25 : 0.04 :
5	0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Excellent	0.00 : 0.03 : 0.31 : 0.37 : 0.29 → Middle	0.00 : 0.20 : 0.44 : 0.02 : 0.33 → Good	0.03 : 0.06 : 0.58 : 0.08 : 0.24 → Good	0.01 : 0.04 : 0.86 :
6	1.00 : 0.00 : 0.00 : 0.00 : 0.00 → Bad	0.64 : 0.00 : 0.07 : 0.28 : 0.01 → Bad	0.92 : 0.00 : 0.00 : 0.08 : 0.00 → Bad	0.87 : 0.03 : 0.02 : 0.06 : 0.02 → Bad	0.86 : 0.00 : 0.00 :
7	0.00 : 0.00 : 0.00 : 0.00 : 1.00 → Very Good	0.00 : 0.00 : 0.00 : 0.20 : 0.79 → Very Good	0.10 : 0.15 : 0.00 : 0.00 : 0.75 → Very Good	0.02 : 0.20 : 0.04 : 0.03 : 0.71 → Very Good	0.01 : 0.60 : 0.00 :
8	0.00 : 1.00 : 0.00 : 0.00 : 0.00 → Excellent	0.00 : 0.80 : 0.01 : 0.01 : 0.18 → Excellent	0.00 : 1.00 : 0.00 : 0.00 : 0.00 → Excellent	0.02 : 0.87 : 0.03 : 0.03 : 0.06 → Excellent	0.00 : 1.00 : 0.00 :
9	1.00 : 0.00 : 0.00 : 0.00 : 0.00 → Bad	0.11 : 0.00 : 0.01 : 0.08 : 0.79 → Very Good	0.47 : 0.25 : 0.14 : 0.13 : 0.00 → Bad	0.29 : 0.09 : 0.27 : 0.25 : 0.10 → Good	0.46 : 0.16 : 0.33 :

Figure 8. Prediction results by different models

Step 5. Evaluating the performance of models on data.

The workflow connects each of the created models with a “*Test and Score*” tool. Once the models are evaluated, it should be seen if their accuracy can be improved by tuning the parameters present in the model.

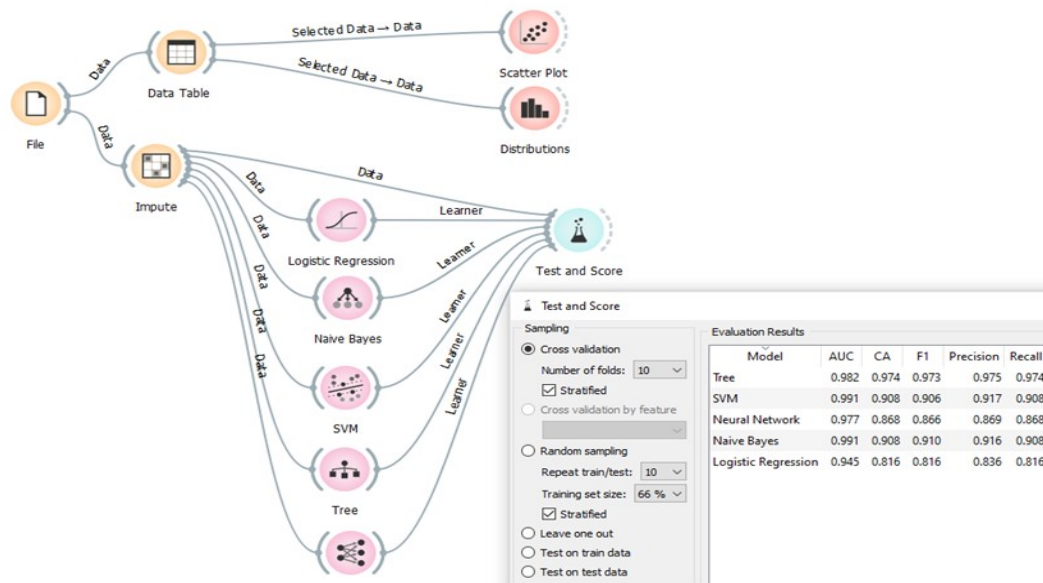


Figure 9: Scoring workflow with the “*Test and Score*” tool

The result of the “*Test and Score*” tool is a table of scores for: *Accuracy*, *Precision*, *Recall* and *F1 Score* for the created models. Specific scores for the quality assessment metrics for the created models in the experiment are given in Fig. 10.

Scores					
Method	AUC	CA	F1	Precision	Recall
Tree	0.965	0.855	0.884	0.864	0.905
Logistic Regression	1.000	0.776	0.976	1.000	0.952
Random Forest	1.000	0.895	0.930	0.909	0.952
SVM	0.975	0.842	0.900	0.947	0.857
Naive Bayes	0.994	0.816	0.952	0.952	0.952
Neural Network	0.941	0.803	0.864	0.826	0.905

Figure 10: *Test&Score* tool operation on the created models

Looking at the predictive accuracy for each of the rating classes, it can be summarized that it is worst for the *Middle rating class*. The highest accuracy is achieved for the *Bad* and *Excellent classes*. *Good* and *Very Good* score predictions for all models considered are presented with an accuracy of around 60-75%. The *Random Forest* model is the most reliable because it performs with the highest accuracy for all classes of estimates. The *Naive Bayes* model has the lowest accuracy score compared to the other models on the data considered.

Testing of the various models and in visualization of the accuracy of the models can be done by "*Confusion Matrix*", *ROC Analysis* or other tool of the system. The workflow is shown in Fig.11.

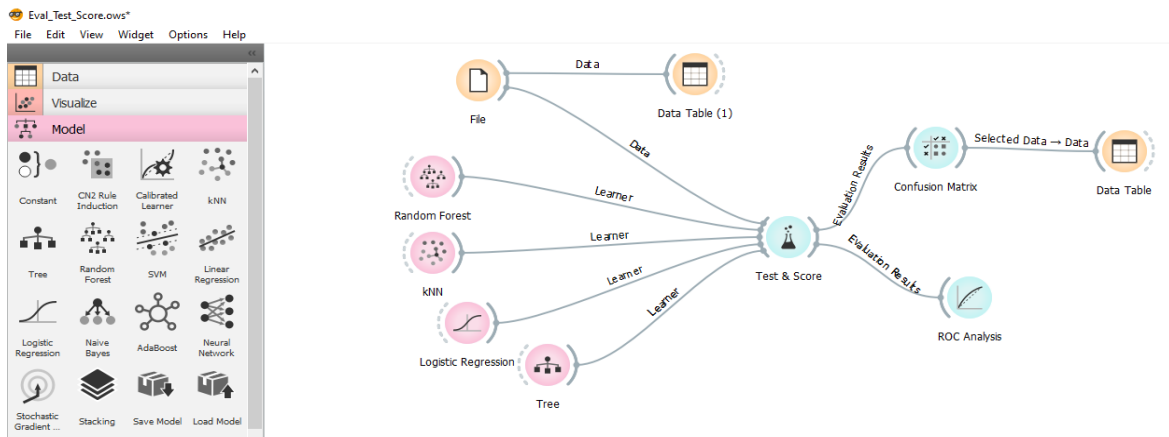
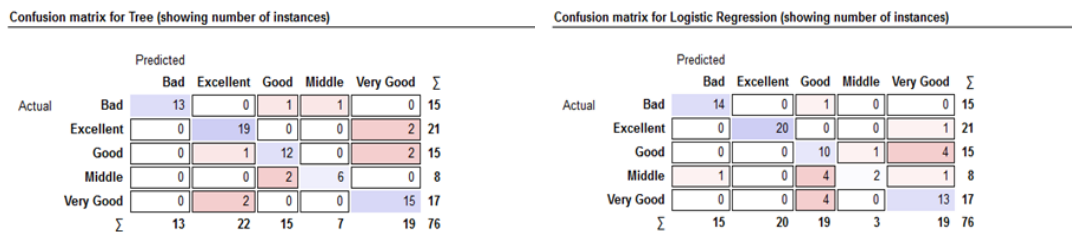


Figure 11. Model creation and evaluation workflow using *Confusion Matrix* and *ROC Analysis*

Results of the work of the *Confusion matrix tool* are given in Fig. 12, to compare the performance of the created *Tree* and *Logistic Regression* models.



Confusion matrix tool performance results for two models created

The *Receiver Operating Characteristics (ROC)* curve compares TPR (True Positive Rate) and FPR (False Positive Rate), which provides information on how well the model correctly recognizes the respective classes. In Fig. 13 the result of comparing four different classifiers is given. The closer the ROC curve is to the upper left corner, the higher the quality of the classifier. The graph shows that in the particular prediction case, the model using the *Random Forest algorithm* performs best.

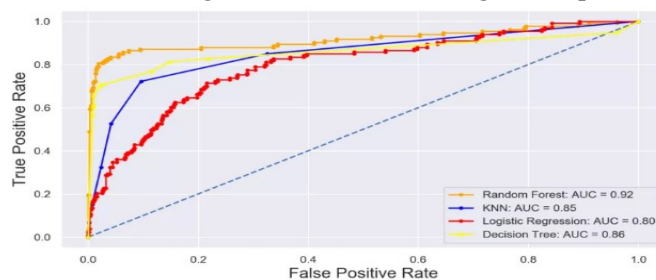


Figure 13. Visualization with the tool *ROC Analysis* in *Orange*

By summarizing the results for the average rating for the predictive accuracy on the rating classes (Poor, Average, Good, Very Good and Excellent) we get: *Decision Tree* - 85.5%; *Logistic Regression* - 77.6%; *Random Forest* - 89.5%; *Support Vector Machine* - 84.2%; *Naive Bayes* - 81.6%; *Neural Network* 80.3.1%. *Decision Tree* and *Random Forest* classification algorithms can be suitable for use in similar evaluation tasks, which predict with a high degree of accuracy the elements of the *Bad* class. These are the students with poor grades that we consider " *students-at-risk*".

In the conducted experiment, the tool processes the data that is collected during the students' learning in their courses. In addition, survey data is also used. The survey is sent by email during the third week of the course. The basis of the survey are questions with which students give their opinion about the tasks, materials and the level of difficulty of the subject. The data obtained from the surveys add a number of new characteristics that are directly related to the dropout of learners, such as lack of interest,

lack of time, organizational obstacles during training, etc. The first experiments were made from a small sample of data, limited by initially available real student data. Using the developed system for monitoring and collecting data on learners and their activity (solved tasks, viewing electronic text and video materials, participation in forums and/or interest groups, etc.) from 4580 records for learners, 15 results from learners are identified at risk. It is a multi-step procedure that is directly related to Big Data Analytics [101] in the e-learning space. A huge amount of data is collected every day, which now requires a new approach to data processing, applying an open-source Hadoop distributed processing framework based on the Map/Reduce algorithm.

CONCLUSION: The conducted empirical studies confirm that the proposed **five-step evaluation method is promising** for the development of an early warning system for various stakeholders of the learning process.

Based on accumulated data from the operation of an e-learning system with different users, applying tools from the field of data science, different decisions about training can be made. The analysis of data accumulated in e-learning courses makes it possible to change the test model and design modules to meet the individual needs of the learner.

Reinforcement Learning (Confirmation Learning)

Reinforcement learning also belongs to machine learning algorithms. The characteristic of this type of algorithms is that it imitates a psychological model, in which the system is given "rewarding" and "punishing" signals in order to maximize the probability of receiving "rewards" and minimize "punishments" [88]. This training is typically applied in the absence of a predefined "correct" training data set. Such an approach is different from the supervised learning approach, where the goal is to reduce the deviation according to pre-set correct data (input/output).

Unsupervised learning

Algorithms of this type take a dataset containing only input values and find a structure or distribution in the data, with no indication of a known variable or reward function, the data is unlabeled, there are no training examples [100]. Main types of solved tasks are: Dimensionality reduction, Density estimation, Clustering.

Cluster analysis is the distribution of a set of observations into subsets (clusters), so that the observations in the same cluster are similar according to one or more predefined criteria, and the observations from different clusters are different [46]. Different clustering techniques work with different assumptions about the structure of the data, which are often defined by some similarity metric and evaluated, for example, by internal compactness, or closeness between members of the same cluster, and difference between clusters.

The clustering methods developed use a different induction principle. Farley and Raftery (1998) [32] propose the division of clustering methods into two main groups: hierarchical and partitioning methods. Han and Kamber (2001) [45] propose a categorization of the methods into an additional three main categories: density-based, model-based, and network-based methods. An alternative categorization based on the induction principle of different grouping methods is presented in Estivill-Castro (2000) [29].

- **Clustering Algorithms**

These algorithms generate different partitions and then evaluate them according to some criteria. They are called non-hierarchical because each instance is placed in exactly one of k mutually exclusive clusters. The desired number of clusters is required to be entered in advance.

One of the most commonly used clustering algorithms is *k-means clustering* algorithms. This type of algorithm belongs to *Exclusive Clustering* because the data is grouped so that if a certain data belongs to a given cluster, it cannot be included in another cluster. Another type of *Overlapping Clustering*, uses fuzzy sets to group data so that each point can belong to two or more clusters with different degrees of membership.

Three-step algorithm: determines the central coordinate (centroid); determines the distance from each object to the center; groups the objects based on the smallest distance. Terminates when the centroids stop moving or some threshold is reached (e.g. number of iterations). The following is an example workflow for applying a k-means tool to a data set and visualizing the result using a Scatter Plot tool (Fig. 14).

To build the information flow, **data collection is done first**. The desktop and toolset are loaded and the "File" widget loads the data for analysis. An *Impute* tool can also be applied to clean the data. Next is **Select Tool k-means**. This *Orange* system tool implements the k-means clustering algorithm. The last step is **to visualize the result**. The tool itself *k-means does not visualize a result, for this purpose a visualization tool, for example Scatter Plot, must be connected to the data stream*.

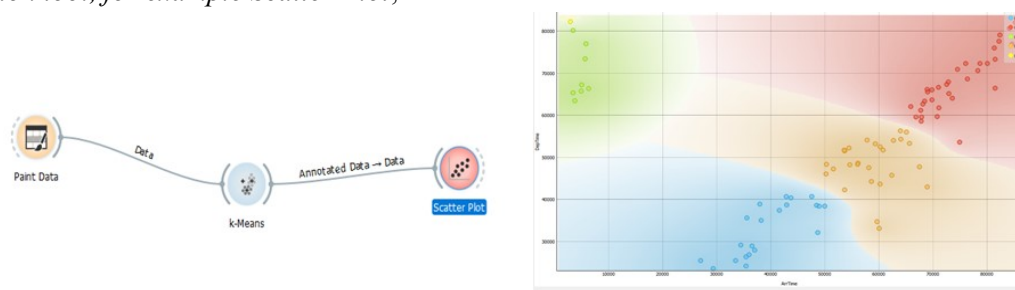


Figure 14. Workflow using *k-means* and *Scatter Plot* tools

- Hierarchical clusterization

When the number of clusters is not predetermined, hierarchical clustering procedures (*Hierarchical Cluster*) are used. These algorithms start by declaring each point as its own cluster and then merge two most similar clusters until a stopping criterion is satisfied.

The wide variety of procedures arises from **the metric used** between different objects. Hierarchical clustering calculates a hierarchical clustering of arbitrary object types from the distance matrix between them and displays the corresponding *dendrogram*. It starts by assigning each element to a cluster. Next, the closest pair of clusters is found and merged into one cluster. Next is a calculation of distances (similarities) between the new cluster and each of the old clusters. The steps are repeated until all elements are grouped into one cluster of size N.

An algorithm for building an information flow in hierarchical clustering is related to the application of the following *steps* :

1. **Data Collection**. Loads the desktop and toolset and loads the data for analysis. A data cleaning tool can also be applied.

2. **Selecting the Distance tool** to calculate the distance between elements. Different geometric metrics can be used to group the data to calculate distance. A *Euclidean metric* is a measure of the distance between points plotted on the Euclidean plane. *Manhattan metric* is a measure where distance is calculated as the sum of the absolute value of the differences between two points plotted on the Cartesian coordinate system. *The Minkowski distance metric* is a summary of the metrics from *the Euclidean metric* and *the Manhattan metric*.

3. **Selection of Hierarchical Clustering tool**. This tool implements an algorithm for the hierarchical clustering of arbitrary object types by the calculated distances and displays the corresponding

dendrogram. A *dendrogram* is a graph-tree in which each node represents one step of the clustering process.

4. **Visualization of the result.** Through different tools, the result can be visualized in different ways, for example through the *Data Table* and *Scatter Plot* tools .

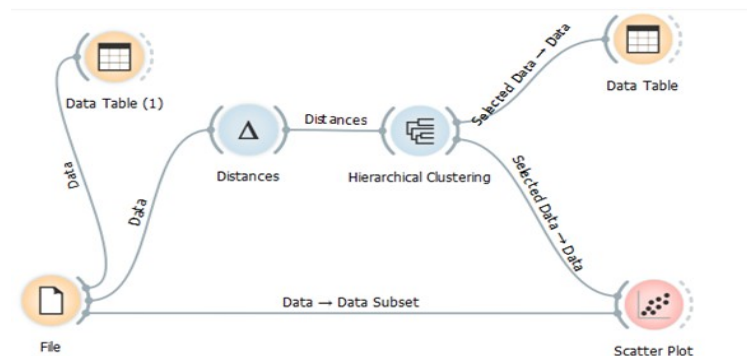


Figure 1 5 . Hierarchical clustering workflow

Other popular knowledge mining software tools that can be used for data analysis are for example: WEKA [103], RapidMiner [87]. KNIME [52], KEEL (<http://sci2s.ugr.es/keel/>), SPSS <http://www.ibm.com/analytics/us/en/technology/spss/> and others.

3.3. Web metrics for evaluating the activities of learners in a virtual educational space

The analysis of the data accumulated during the training allows research and feedback on how the learners search for information, what difficulties they encounter and the design of modules that meet the individual needs of the learner. A new element here is an assessment of the degree of use of websites related to the field of learning and modeling of learners, based on an analysis of their behavior in the learning environment and the web space.

Web analytics is the measurement, collection, analysis, and reporting of Internet data in order to better understand the complex interactions between website users. The web analytics process includes various web metrics defined in Google Analytics [25] as:

- **page views** - the number of views of the web page accessible by a visitor (without spiders or robots);
- **visitors** - the number of unique visitors to the website;
- **pages/visits** - the number of pages viewed by a visitor during a visit;
- **time on site** - length of time spent by all visitors to the website;
- **stickiness** - the ability of the web page to keep the visitor on the web site;
- **frequency** - number of visits by a visitor to the site (loyalty indicator);
- **recency** - number of days that have passed since the visitor's last visit to the site;
- **length of visit** - the visit time spent by a visitor on the website (in seconds);
- **depth of visit** - number of pages visited by a visitor in one visit, etc.

In the field of web analytics, there are mainly two techniques used to analyze website traffic: server-side and client-side data collection. Server data collection methods extract and analyze data primarily from log files and include log information such as IP address, time and date, browser type, etc. Behavioral information includes general browsing information such as number of pages viewed, language setting, etc. When data is collected from client sites or bookmarked pages, the page visitor data is sent to a tracking server using a JavaScript code (or tag) inserted into the HTML page. With this approach, all the actions of the visitor can be accurately tracked, as well as additional information can be collected. Cookies can be used to determine how many first-time or repeat visitors a site has received, how many times a visitor returns in each period, and how much time passes between visits.

Google, WebTrends, Nedstat [5] and many other companies provide web analytics software using page markup. Google Analytics is the most used free program [25]. From the point of view of the analysis of the activity of the learners, the information related to the customers, the history of their visits and behavior, the user profile, etc., is particularly interesting. Google Analytics prepares anonymous and statistical reports about the websites that use it. Show data such as geographic location (based on general IP-based geolocation codes), time of visit, etc. For example: the report in Fig. 16 [25] makes it possible to track the site's visitor traffic by hours.

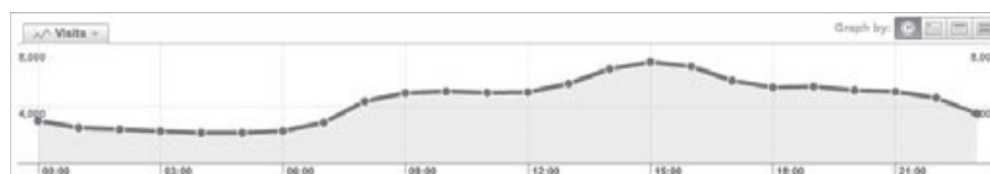


Figure 16. Graph aa hourly trend with Google Analytics.

In order to analyze the web sites most visited by the learner, the depth and frequency of visits, *the Fuzzy Classification of the Web Metric* [107] is used. Fuzzy set theory and fuzzy logic are known to account for the imprecision, uncertainty, and ambiguity of human thought and language by defining the membership function.

A fuzzy set A in X is defined as [105] :

$$(9) \quad A = \{(x, \mu_A(x))\}, \text{ where } x \in X, \mu_A : X \rightarrow [0, 1]$$

is the membership function of A and $\mu_A(x) \in [0, 1]$ is the degree of membership of element x in set A. In our experiments on the number of visits to a given website, the terms "low", "medium" and "high" as language variables. In general, reporting pageview values as true (1) or false (0) for a month, pages are defined as "low visited" if visits are eg between 0 and 25, between 25 and 55 visits pages are "average visited" and more than 56 visits are classified as "high view". However, if a visitor has 55 page visits, they are classified as "moderately" active, while another visitor with 58 visits is classified as a "very" active visitor. Despite a difference of only 3 visits, they are divided into two different groups. By defining fuzzy sets represented by a membership function, a continuous transition between the classes "low", "medium" and "high" is introduced. Thus, a visitor can belong partly to two classes (55% of "highly active" and 45% of "medium active") at the same time. The use of fuzzy classes allows a more precise classification of web metric values [50]. Taking this into account in order to estimate website traffic, a fuzzy rule system is developed [28]. The pageview measurement is a number that doesn't have much meaning by itself. Only the context of the number against the number of other pageviews provides knowledge on which to make an estimate.

Fuzzy classification rules are implemented to determine the number of web page hits. The method of *inductive fuzzy classification* (*Inductive Fuzzy Classification IFC*) is applied, in which the grouping of the elements in a fuzzy set is performed with a membership function inferred by data induction. Inductive Fuzzy Classification by Percentile Rank (IFC-PR) [50] generates a fuzzy membership function, in this case using the common language terms "low", "medium" and "high" correlating with the number of user visits:

- the empirical rank of the value x with the metric M determines the belonging to class "**high**":

$$\mu_{\text{high}}(x) := P(M < x)$$

- the indicator M will be classified as "**low**" (the negation of belonging to the "high" class):

$$\mu_{\text{low}}(x) := 1 - \mu_{\text{high}}(x)$$

- "**medium**" classification is defined as:

$$\mu_{\text{medium}}(x) := 1 - \text{abs}(\mu_{\text{high}}(x) - 0.5) - \text{abs}(\mu_{\text{low}}(x) - 0.5).$$

In the conducted research [82], an analysis is made of the visits to the websites indicated as supporting materials in the course on the subject of Artificial Intelligence. For each web page, the visits are estimated, for example: page W1 has 135 visits within the training course. Total suggested and monitored pages in the course are 80 and 56 of them have less visits than W1. To estimate W1 visits, calculate:

$$\mu_{\text{high}}(\text{visit}(W1)) :=$$

$$P(\text{Number of visits} < \text{visits}(W1)) = 56/80 = 0.7.$$

Thus, according to the fuzzy classification, page traffic is defined as "high" above 0.7, "low" below 0.3, and "medium" up to 0.6.

Based on the resulting analysis, for each web page offered in the training course, decisions are made to update the web sources, as well as supplement the literature with new sites related to the most sought-after topics.

After analyzing the web sites most visited by the learner, the depth and frequency of visits, the type of their content should be analyzed and determined. An approach to automatically determine the type of document and how relevant it is to the learning domain is presented in [80]. The analysis is based on the idea: certain words in the document are relevant to its content. To determine these words, the frequency of their occurrence is searched using the *Wordcloud software* [104]. Filters can be applied on the received word list. The proximity of each word from the resulting set to all words from the subject

ontology dictionary associated with the learning domain is calculated and the smallest value obtained for each word is taken. To determine the proximity of words and phrases, an approach presented in [48] using q-gram metrics is used.

By processing a large set of web-documents, a dataset is accumulated with central words extracted from them and their degree of proximity to the concepts of the analyzed domain. These data were used to train classification algorithms, presented and described in the form of a multi-step process, with the target attribute being whether the document is related to the domain or not. An algorithm for the analysis and prediction of documents from the considered field of study is proposed, shown with a diagram in Fig. 17.

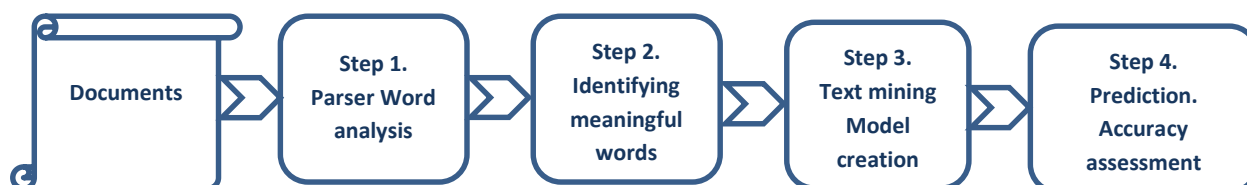


Figure 17. Multi-step process of analysis and prediction of documents from a given domain

From the collected information about the activity of users in the web space by means of cluster analysis or application of associative analysis, user segmentation can be performed. Web content can be analyzed in the following steps (Fig. 18).

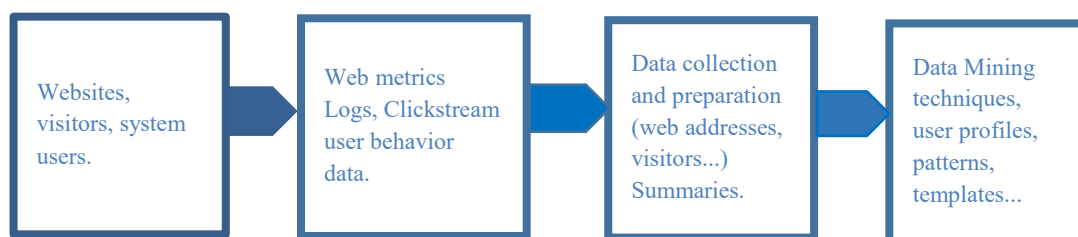


Figure 18. Web metrics data extraction architecture

CONCLUSION: In the stage of analysis of the results of the conducted training, special attention is paid to the difficulties that the trainees encounter when completing the individual modules in the discipline. In the conducted experiment, the conclusions drawn are related to the fact that students in the studied discipline of Artificial Intelligence encounter difficulties when working with the literary sources for preparing the project on a selected topic from Module 4, in the areas: "Blockchain", "*Ethics and Emerging Sciences*", "*Policy and investment recommendations for trustworthy Artificial Intelligence*" as well as in determining risks related to these technologies.

Through the presented method in the learning environment, it is possible to track the activity of learners in the Internet space, accumulating data regarding their activity in the learning environment. By analyzing this type of data, timely information is obtained about the activity, progress and success of learners. The results show that the formulated solutions can be successfully used for different tasks and can be adapted to new technologies and applications. The proposed method promotes the creation of an innovative learning environment and is the next step in the digitization of education.

3. 4. Modeling the user in the educational space

A user model is built through appropriate means of monitoring and reporting the user's activity. The model serves to customize the system to the knowledge and skills of the user and apply adequate supporting strategies in his work. **The user model** is essentially a data structure representing individual user characteristics of a given program system. It represents the user's cognitive processes and perceptions of the application domain [7].

According to Barr and Feigenbaum [10], there are two **basic approaches**:

- In *the overlay* approach, the user's knowledge is represented as a subset of the general knowledge maintained by the system for the domain being studied, the domain expert's knowledge, or the expected knowledge of the learner. The knowledge of the system is decomposed into independent components and covered with a system of notations indicating the level of mastery of each individual component. Accepting this opinion, a model is built here for assessing the basic knowledge of the learners regarding the main concepts and dependencies between them. A hierarchy of concepts defined in the used domain-ontology is built in the taught subject area. After each test, the concepts are assigned a relative numerical value (in percentages), which indicates the degree of certainty of the system regarding the knowledge of this concept by the specific learner. The rating for each concept can be formed dynamically as an averaged rating of its child concepts according to the formula:

$$(10) \quad \text{Mark}_{\text{term}} = \frac{1}{k} \sum_{i=1}^k \text{Mark}_{\text{Subterm}[i]},$$

where $\text{Mark}_{\text{Subterm}[i]}$ is the evaluation of the i th child concept, and k is the number of child (inherited) concepts of the evaluated main concept. Each child concept, in turn, can be considered as the parent of its child concepts and receive its evaluation according to the same formula. Then the general assessment for the taught topic can be formed according to the formula:

$$(11) \quad \text{Mark} = \frac{1}{n} \sum_{j=1}^n \text{Mark}_{\text{Term}[j]},$$

where $\text{Mark}_{\text{Term}[j]}$ is the j -th score main concept in the studied topic of the subject area, and n is the number of these main concepts.

- In the *bug theory modeling* approach, bugs are formally diagnosed through a list of pre-defined incorrectly learned and missing elements of knowledge. Here, the user's model consists of the expert's model, supplemented by a list of domain errors. Much of this data is collected from the e-learning environment and stored in a relational database. Using the accumulated data, reports on the participants and their work are generated, for example: a report on the participants in a given activity, a report on the duration of a given training, the start date and the completion date of the participation; data on learners and the courses they participate in; total number of participants; number of participants trained in individual courses, etc.

In order for learning to be formally represented at the meta-level, its basic aspects (knowledge, processes and participants) must be modeled and formalized in a way that ensures defining, finding and using relationships between them. The dissertation develops and proposes a model of the learning process in the educational space, based on ontologies for knowledge representation. Figure 19 presents a diagram of the developed ontological model of an electronic course. A prototype of the ontology was tested using the Protege system (<http://protege.stanford.edu/>).

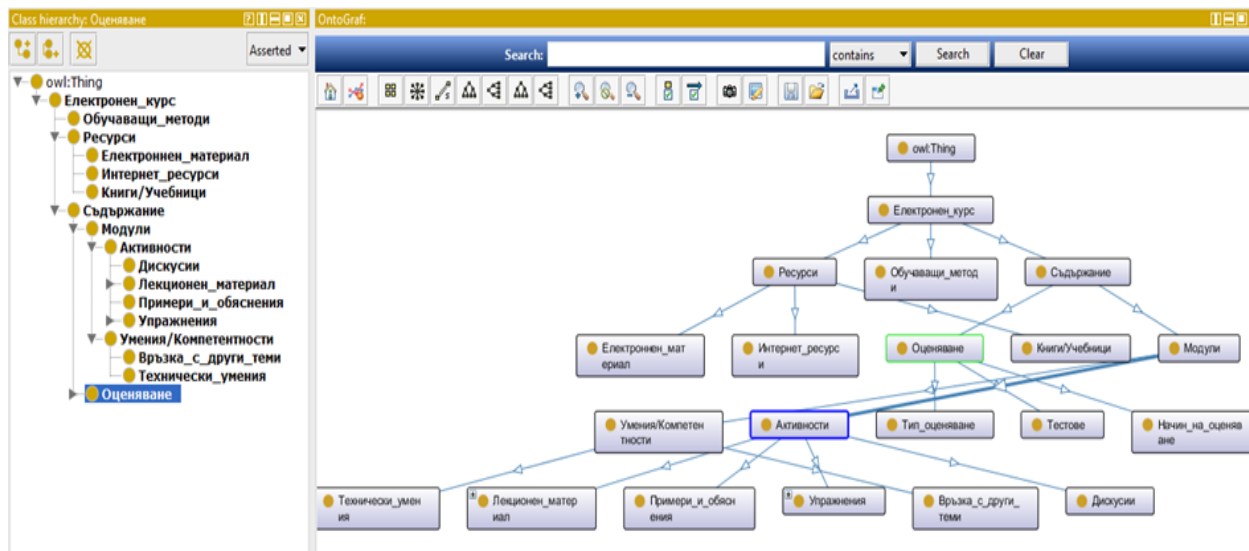


Fig. 19. Ontological model of an electronic course through the Protege system

The learning domain knowledge model represents the core concepts and their relationships that must be taught and tracked as a level of mastery.

Based on experiments with learners in an electronic environment in various disciplines, a general framework of a learner model architecture is proposed here, which includes three types of factors: competence factors, emotional factors, social environment impact factors.

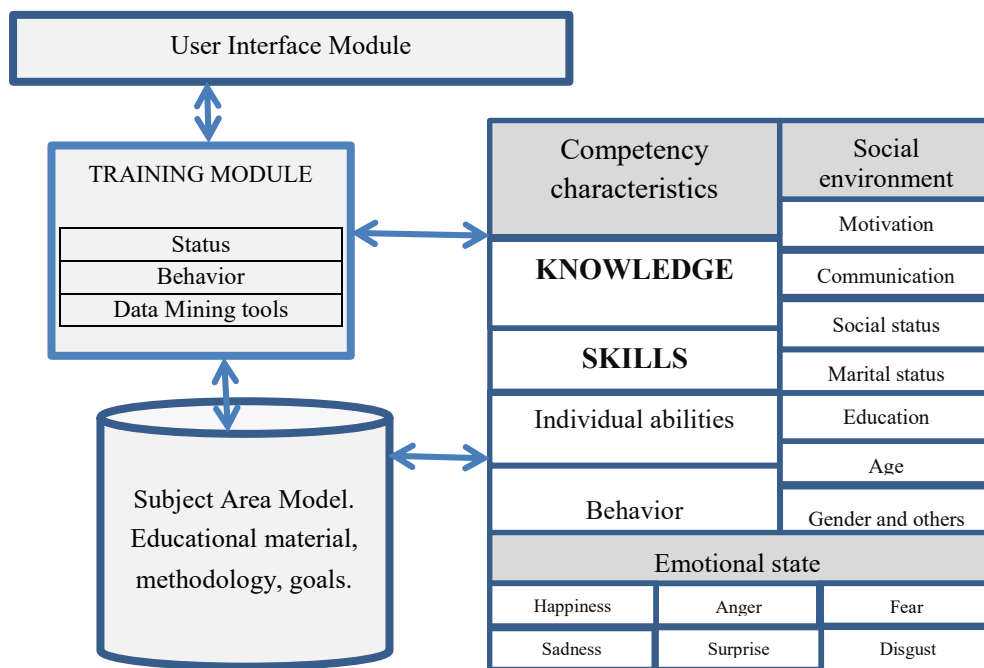


Fig. 20. Architecture of a learning system with learner model

The proposed model of the user is distinguished as a separate module containing user data and a system of functions taking care of the collection, storage, processing and interpretation of this data. This module is related to the model of the subject area of study and the learning module of the system. The individual model (Fig. 20) contains three types of data.

- The first type presents assessments on elementary skills and typical errors in work in the field of study. These evaluations are summarized in the overall evaluation of the learner in the area.
- The second type shows the level of assimilation by the learner of each of the units of knowledge. For a learner, a list of elements of the type: <units number> : <evaluation of units> is maintained.

- The third type concerns the learner's work during the training session: it contains the curriculum number, the current difficulty of the questions and the tasks; the level of assistance, etc. This also includes data about the learner such as: gender, age, cultural and social status (includes forum posts and number of responses, group interaction), as well as some individual characteristics. This component is in the process of experimental research and the difficulties stem from the fact that no quantitative theory is known about the relationships between the psychological characteristics of the learner.

Ratings indicating the level of mastery or non-mastery of a given knowledge or the ratings of questions, tasks or requests of the learner are real numbers in the interval $[-1,1]$.

Combining this approach with the approach of maintaining an error library, for each of the predefined error types, counters are maintained for the times when an error was possible and the number of incorrect attempts. After the completion of the event, for each type of error, an estimate is formed:

$$(12) \quad \text{Mark}_{\text{current}} = 1 - \text{Count}_{\text{error}} / \text{Count}$$

Here, $\text{Mark}_{\text{current}}$ is an intermediate score that is a number in the interval $[0,1]$ and actually represents the percentage of attempts by the user that did not make an error. Count is the total number of attempts, a $\text{Count}_{\text{error}}$ is the number of wrong attempts. To get the final Mark score for the skill in the interval $[-1,1]$, it is calculated using the formula:

$$(13) \quad \text{Mark} = 2 \cdot \text{Mark}_{\text{current}} - 1$$

Taking into account the previous two formulas, for the estimate we get:

$$(14) \quad \text{Mark} = 1 - 2 \cdot \text{Count}_{\text{error}} / \text{Count}$$

In the case where the total number of attempts is 0, we assign the score a value of 0. Once a post-event error type score is formed, it is reflected in the score for the error type in the user model by the formula:

$$(15) \quad \text{Mark}_{\text{new}} = (1 - K) \cdot \text{Mark} + K \cdot \text{Assessment}_{\text{of}_{\text{the}}_{\text{assignment}}}$$

Here, K is a number in the interval $[0,1]$ and represents the influence that the error type estimate has on the estimate in the learner model. It is calculated according to the formula:

$$(16) \quad K = 0.5 \cdot (\text{Count} - 1) / \text{Count}$$

Thus, the assessment for each of the skills contains information from previous tasks and from the last task. The score is multiplied by 0.5 to limit the unwanted influence that the old score has on the new score.

The calculation of second-type assessments for the level of mastery (or lack of mastery) of individual units of knowledge is carried out as, the initial value is assigned 0, and after each solved task or question, the grade is changed according to the formula:

$$(17) \quad \text{Evaluation}_{\text{new}} = (1 - K) \cdot \text{Rating}_{\text{old}} + K \cdot \text{Problem}_{\text{Assessment}}$$

but here $\text{problem}_{\text{grade}}$ is the grade of the question or task given by the teacher. K is the influence that the evaluation of the problem has on the evaluation of the unit of knowledge and is calculated by the formula:

$$(18) \quad K = \text{Problem}_{\text{Difficulty}} / \text{Unit}_{\text{Difficulty}}$$

the greater the difficulty of the problem (question/task), the greater its influence on the evaluation.

Scores from the interval $[-1,1]$ are interpreted as follows. The closer the score is to 1, the more certain the system is that the learner has correctly mastered the given unit of knowledge. The closer the score is to -1, the more certain the system is that the learner has incorrectly mastered the given unit of knowledge. And the closer the score is to zero, the more the system is not sure of the learner's knowledge. Such an assessment approach takes into account characteristics of both the "overlap modeling" and "error library modeling" approaches. The following is meant here: if the assessment is positive, it represents the level of the learner's knowledge compared to the expert's knowledge (typical of the first approach), if the assessment is negative, it represents wrongly acquired knowledge (as in the second approach). Based on these considerations, it can be argued that the presented approach is of a hybrid type.

And they also use counters that monitor the number of used educational resources and the type of Internet resources visited, which in different combinations can provide a different presentation of the educational materials. Information resources are evaluated by number, types, etc.

Emotional state recognition in the user model in an e-learning environment

For the purposes of the research conducted in the dissertation, a software application was developed for recognizing human emotions based on the image of the face. The developed software is based on the facial action coding system (FACS) of Ekman and Friesen [27]. The facial expression of emotions is determined by the muscles involved and the gestures that characterize it. There are certain distinctive patterns of emotional reactions generalized and shared by most people. They are considered basic emotions: *happiness, sadness, anger, surprise, fear* and *disgust*. Other similar coding systems are EMFACS, MAX, AFFEX and CANDIDE-3. The developed software and the conducted research are presented in [57].

Methodology. The first step is to categorize the set photos with images of faces into separate groups (subdirectories). They are seven in number (including *Neutral*), and each of them corresponds to a specific emotion and contains about fifty sample images related to the corresponding emotion. A next step is to use the *Dlib* library of algorithms to create general-purpose software that is used in many fields. By using the *spape_predictor_68_face_landmarks.dat* file, the faces from the images are marked with corresponding cardinal points. Based on the set points, in the next step, mathematical calculations are performed, with the help of which 6 normalized vectors of each face are determined. Each facial feature vector was calculated by normalizing the data and summing the distances between the marked points. By means of the "ML.Net" framework, a multi-class classification algorithm based on a Supervised Learning ML agent is implemented.

After training the model, it can recognize the emotional state of a given image (photo) of a face. As a result of the work, the program returns a prediction about the emotional state. The analysis of the results of the metrics for evaluating the accuracy of the model shows that the highest prediction accuracy is achieved for the emotional states *Joy, Surprise* and *Sadness*. *Anger* state is often confused by the algorithm with *Sadness, Joy* or *Neutral*. The algorithm can be improved by training with a much larger amount of training data from different types of emotional states. But a balanced data set should be used.

An alternative approach is to apply an *Unsupervised Learning algorithm*, which examines the structure of the data without instructions, the training data being unlabeled. *Clustering* algorithms or *Deep Learning* can be used. By applying different classifiers one can experiment with the obtained results to increase the accuracy, against different groups and categories of data. The goal is to integrate the developed application into an e-Learning environment and to look for changes in the emotional states of learners while they are dealing with cognitive tasks in the learning process.

Assistive device for hearing impaired users

A device was designed and experimentally investigated [106] that produces a wide range of colors and intensity levels to visualize different aspects of sound. For the hearing impaired, sound visualization provides an enhanced sense of the environment and allows them to interpret and respond to auditory information in real time.

Software tools have been developed for the analysis of sound frequencies and their conversion into the corresponding colors in the RGB model. A model of an automated system was created, which realizes the conversion of sound frequencies into color. The achieved accuracy has been evaluated experimentally. It was found that the accuracy of the conversion depends on the frequency of the sound. Error rates have been shown to be higher at high sound wave frequencies (above 1000 Hz). Work should continue with research aimed at generalizing the models for the conversion of sound frequencies into color.

3. 5. A cognitive agent in an e-learning environment

A cognitive agent [24] is autonomous software that has the ability to develop its knowledge. Given the varying degrees of intelligence, cognitive agents can be implemented based on rules that develop their knowledge using inference mechanisms built into the system or extract new knowledge using machine learning and data mining techniques.

Rule Based Systems system built for this study includes the components [89]:

- o **Knowledge base** (a set of rules) - contains all the knowledge needed by the agent.
- o **Base of data**: contains the data that is established at the current time.
- o **The Inference Engine** - the interpreter selects and applies the rules that can lead to a change in the state of the database. Different approaches exist for this purpose: forward chaining, backward chaining and mixed approaches [89].

o **The base of meta-knowledge** - meta-rules defining the relationship between the rules [24]. The way in which the rules are selected from the set of applicable rules is the strategy for solving conflicts (Conflict Resolution). In this case, rule priorities are applied, which are dynamic. Dynamic priorities are determined based on the importance of actions in the current situation. For example: if *the learner's score is high*, then a low priority is set to the rule to search for the next task to solve. But if it has *a low score*, the rule for solving the next task should have a high priority. The learner has to solve the next problem to increase his score. The approach is flexible but time-consuming.

On the other hand, **Reinforcement Learning** is becoming an increasingly popular type of machine learning. It is a technique where autonomous agents use trial-and-error algorithms and a cumulative learning reward function. The advantage lies in the calculation of optimal actions that agents can take within scenarios determined by the environment. This training is typically applied in the absence of a predefined "correct" training data set. Such an approach is fundamentally different from the way of operating in supervised learning, where the goal is to reduce the deviation according to pre-set correct data (input/output). Training proceeds as a sequence of trial actions that gradually lead to the reinforcement of good actions and the avoidance of inappropriate ones. The result of the training is an optimal strategy for action in any situation. A strategy is optimal if it manages to maximize the sum of all rewards received during its execution.

Formally learning through Reinforcement Learning is considered as [54] :

- A set S including all states of the environment that the agent recognizes.
- A set A , including all the actions that the agent can perform.
- A set of rules for transitions between states.
- Rules that determine the R - rewards that an agent can receive in transitions.
- Rules that describe what the agent abides by.

A specific task set in the dissertation is to analyze and compare the behavior of two agents implemented on the basis of two different approaches based on machine learning. The results of the conducted research and the software developed for the purpose are presented in [58]. For analysis purposes, a game was created that provides a platform and generates x number of bombs and coins on it. The game creates two agents represented as two different balls on the platform - red and blue. The platform has no surrounding walls, which means that the balls (agents) can fall. If an agent falls, they lose the game. Each agent can collect coins, with each coin earning the agent a point. If an agent touches a bomb, they lose a point. The game ends when the coins run out or when an agent falls off the platform. The agent with more points wins the game.

An agent, represented by a blue ball in the game, maintains behavior based on a system of rules created by the programmer and built into the system's knowledge base. In this case, the *Rule-Based System* approach is applied when building the game. On the other hand, an agent represented by a red ball

is implemented based on the machine learning approach with *Reinforcement Learning*. With it, the artificial intelligence learns by itself what is the best action in each situation and optimizes the decisions it makes over time.

The differences in the behavior of the solutions created by the two chosen approaches are examined. The hypothesis is that an agent implemented with machine learning through *Reinforcement Learning* will have more diverse behavior and, after sufficient training time, will perform more efficiently than an agent implemented as a rule-based system.

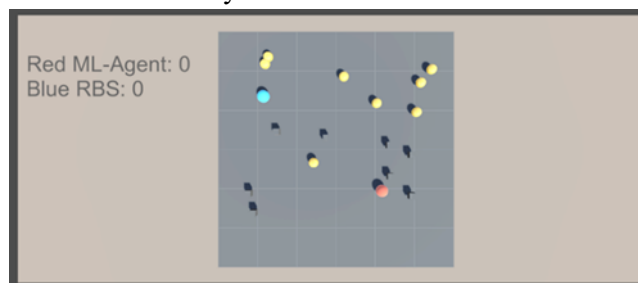


Figure 21. View of the developed software

The reason for choosing *Reinforcement Learning* is the wide use of this approach in the gaming industry [54]. More experimentation and understanding of this type of machine learning is helping to expand the scope of its use in modern learning environments. On the other hand, the alternative to rule-based systems is an approach that has time-proven its advantages and effectiveness for quickly achieving the set goals. The project [58] was created using the *Unity platform*, along with *Unity Technology's ML-Agents* machine learning agent development suite. This package provides a variety of training and validation scenarios. Each agent can have a set of states and observations, take actions in the environment, and receive rewards for events in it.

- *Research methodology.*

The first Reinforcement Learning agent was created by inheriting from the *Agent class*. Redefined the *OnEpisodeBegin method*, which sets the initial parameters for the game (coins and bombs and their positions). Redefining the *CollectObservations method* provides passing current direction information to the nearest coin. *OnActionReceived method* sets the allowed movements of the agent. A *Ray Perception Sensor* component allows the agent to observe objects in the game world by receiving information about coins and bombs in the area and observing where the platform ends. *ML_agent* is the main class in this project, it determines where coins and bombs are generated, keeps track of the score and determines the winner at the end of the game. When an agent touches an object, the action is detected and the "*OnTrigger*" function is called, which allows to see which object the agent came into contact with. The score increases on contact with a coin and decreases on contact with a bomb. The current score screen shows the scores of both agents.

The second agent, based on a rule system, is implemented through *the RBS class*, and the agent's behavior is set by the implemented methods in the class. Depending on the scenarios, the rules apply: "If there are coins on the platform, go get the nearest one"; "If the other agent is near the end of the platform, try to knock him off the platform"; "If the RBS agent has more coins than the other agent, he should not be afraid of the bombs to finish the game faster and not allow the opponent to collect more coins. Rule enforcement is based on the forward chaining approach.

CONCLUSION: *Reinforcement Learning* agent gradually learns what is the best action in any given situation and optimizes the decisions it makes over time. It initially works inefficiently, but over time it will optimize its actions. On the other hand, provided that good enough rules are created, a *Rule-Based System* is a programming task that can support very complex behavior. An interesting trend in more complex applications is the joint application of *Reinforcement Learning* with supervised and unsupervised learning, if the methods used alone do not give a good enough result.

3. 6. Models for hierarchical multi-component assessment of learners

New models for hierarchical multicomponent assessment of learners are presented here , which aim to comprehensively assess various high- and low-order thinking skills, theoretical knowledge and practical skills, etc.

Bloom's taxonomy [13] defines a hierarchy of thinking skills in which higher levels of thinking include all cognitive skills from lower levels. The levels are structured as: *Knowledge, Comprehension, Application, Analysis, Synthesis* and *Evaluation*. Each level is defined by multiple cognitive skills, activities and assessment methods. It is considered that in the learning process, the learner passes through all levels sequentially. A generally accepted classification defines the skills from the upper three levels of Bloom (Analysis, Synthesis and Evaluation) as higher-order thinking skills (HOTS), and those from the lower three levels (Knowledge, Comprehension and Application) - as lower-order thinking skills (LOTS). HOTS includes critical thinking abilities, knowledge and skill transfer skills, problem solving skills, etc. [15]. Student assessment must be objective, and in many cases it is difficult, especially when it comes to assessing HOTS, which require creative thinking. This motivates work in the direction of creating complex multi-component models for evaluating student achievements [36].

A modeling of the hierarchical organization of evaluation components and the dependencies between them, which the evaluator explicitly or implicitly uses in evaluation, is proposed. For example, in multiple disciplines, the main assessment components are 'practice' and 'theory', and in the terms used here: *thinking skills*, their respective sub-components can be theoretical and practical LOTS and HOTS. On the other hand, if we choose HOTS and LOTS as main components, they could have sub-components for assessing theoretical knowledge and practical skills. The creation of a specific hierarchy of components and subcomponents can be done using pre-standardized models selected by the evaluator. Given enough input-output samples and effective neural network training, the results obtained can reflect this hierarchical organization, even without it being explicitly specified.

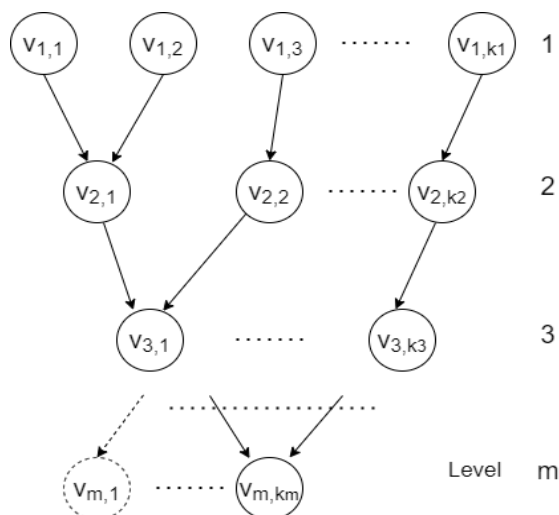


Figure 22. a) **Model 1:** Hierarchical tree-like organization of evaluation components

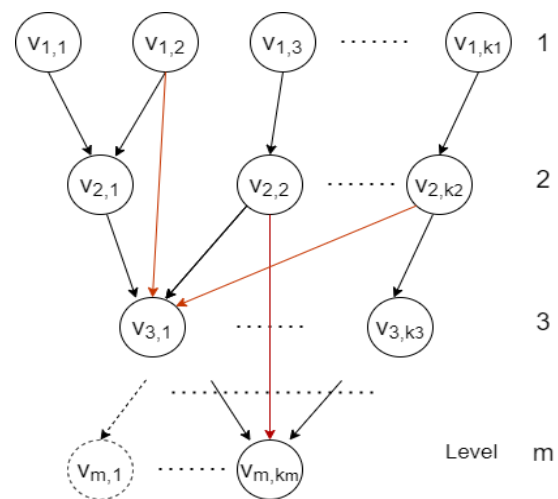


Figure 22. b) **Model 2:** Hierarchical graph organization of evaluation components

In Fig. 22 a) a general **Model 1 of a hierarchical tree-like organization of evaluation components** is presented, in which each main component, in particular the final evaluation, depends on many sub-components of the previous level. In doing so, each sub-component affects only one higher-level component.

Model 2 of a hierarchical graph organization of evaluation components (Fig. 22 b), extends and generalizes the capabilities of Model 1, with the possibility of one component influencing multiple components of arbitrary levels and depending on components of various lower levels.

The presented models can be used in many different scoring approaches where the final score is a function of multiple scoring components. The main characteristics of the proposed models are:

- Level 1 of the component hierarchy describes specific values from conducted evaluations, which can be with different evaluation scales. Examples of such values are: theory and practice test scores; scores from skill tests, grades from assignments, projects, and more.

- At each subsequent level, the components form a score that is a function of the scores of its subordinate sub-components from the previous level. In particular, all or part of the functions may be fuzzy logic.

- At the last level there is one component $k_m=1$, but not required. There may be several final grades, e.g. for different cognitive skills assessed. Thus $k_m \geq 1$.

A possible option for Level 2 features is to normalize all Level 1 component values to a common scoring system (eg 2 to 6). Similarly, the functions at each subsequent level could preserve the normality of the parent components in the corresponding rating scale.

Typically, the evaluative sub-components have values in the space of positive real numbers R^+ . In different evaluation approaches, the values may also be in other real, complex, or other number spaces. The final score E necessarily belongs to a space of predetermined possible values – for example, it must be an integer in the interval from 2 to 6, from 1 to 5, from A to F, etc.

In the formal mathematical description of **Model 1** and **Model 2**, any level with number i owns k_i of number of components defining the set:

$$(19) \quad V_i = (v_{i,1}, v_{i,2}, \dots, v_{i,k_i}) \in R^{k_i}, k_i \in N, i = 1, \dots, m$$

Level 1 components are specific values obtained during evaluation. At each subsequent level $i = 2 \dots m$ values are obtained as a result of applying functions to the components of the previous level. In **Model 1**, the number of components in a given level decreases or is kept equal to the number of components in the previous level $k_s \leq k_{s-1}, s = 2, \dots, m$, which is a consequence of merging sub-components into components. The number of summarizing functions and, accordingly, of the components, can increase in certain levels in **Model 2**.

The values of the components, after level 1, in **Model 1** are calculated as a function of the values of the components of the immediate parent levels:

$$(20) \quad v_{i,j} = f_{i,j}(P_{i-1,j}), P_{i-1,j} = (v_{i-1,s_1}, v_{i-1,s_2}, \dots, v_{i-1,s_j}) \subseteq V_{i-1}, \\ \forall i = 2, \dots, m, j = 1, \dots, k_i, s_j \leq k_{i-1} \text{ and}$$

the components of V_{i-1} occur only once in a while $P_{i-1,j}$.

In **Model 2**, parameters of the functions at a given level can be arbitrary evaluation components from all previous levels, thereby dropping many restrictions:

$$(21) \quad v_{i,j} = f_{i,j}(Q), Q \subseteq \bigcup_{r=1}^{i-1} V_r, \forall i = 2 \dots m, j = 1 \dots k_i.$$

An experiment using fuzzy logic in hierarchical multicomponent scoring

In the experiment conducted in the discipline "Programming on the Internet", through four exam components, we evaluate theoretical knowledge and practical programming skills:

- tLOTS – theoretical LOTS, evaluated by points in interval [0, 30];
- pLOTS – practically LOTS, the score is an integer in the interval [0, 15];
- tHOTS – theoretical HOTS, the estimate is an integer in the interval [0, 15];
- pHOTS – practically HOTS, the score is a real number in the interval [2, 6].

pHOTS assessment is formed by a practical task and is in the interval [2, 6]. The points of the remaining evaluation components are formed when solving a test in a paper or electronic version. The final grade $finalN$ is an integer on the grading scale [2, 6], i.e. $finalN \in [2, 6] \subset N$. Based on **Model 1** of a

hierarchical tree-like organization of evaluation components, we define **two main possible options for formalizing the logic of the evaluator** for assessment in the discipline (F

ig. 23). If necessary, by defining additional connections/dependencies between components at different levels, we can move to **Model 2**. In practice, in different situations, other options can be created and experimented with. What is important in all cases is to determine the hierarchy of components and relationships between them, as well as the set of their corresponding aggregating functions, for the given assessment.

Main components forming the final assessment in **Option 1 for theoretical-practical hierarchical multicomponent assessment** are **theory** and **practice** (Fig. 23 a). Sub-components of the theory are **the normalized forms** of the primary components **tLOTS** and **tHOTS**, and in practice – the normalized forms of **pLOTS** and **pHOTS**. The purpose of the normalized forms is to reduce the primary components to the same rating scale (in this case from 2 to 6), so that they are presented to the evaluator's thinking frameworks for the evaluation.

On the other hand, in **Option 2 for HOTS - LOTS hierarchical multicomponent evaluation** (Fig. 23 b), HOTS and LOTS are adopted as main components, a their sub-components are respectively the normalized forms of the practical and theoretical HOTS, and the practical and theoretical LOTS.

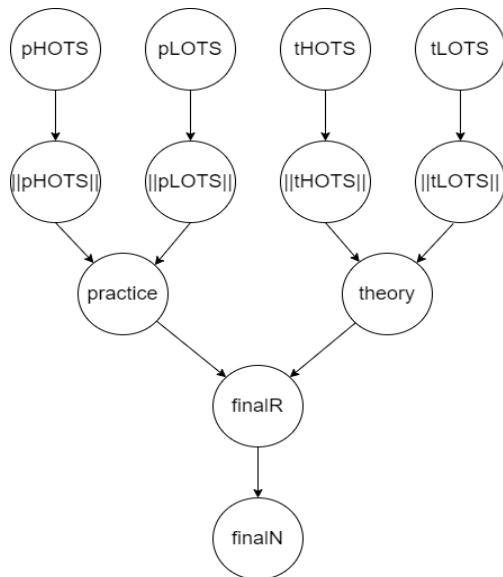


Figure 23.a) **Option 1**: Theoretical-practical hierarchical multi-component assessment for **Model 1**

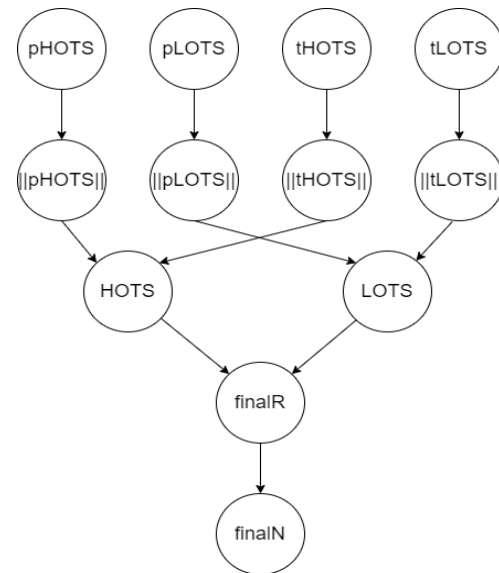


Figure 23, b) **Option 2** : HOTS - LOTS hierarchical multivariate estimation for **Model 1**

In both modeled versions of the first functional level, the forms of the main evaluation components are normalized. Specific normalization functions may follow **a liberal** or **strict** scoring approach, or their modification, depending on the appraiser. In **the liberal approach**, the points of the primary components are reduced to the interval [2, 6]:

$$(22.1) \quad ||tLOTS|| = 2 + 4 \frac{tLOTS}{30}$$

$$(22.2) \quad ||tHOTS|| = 2 + 4 \frac{tHOTS}{15}$$

$$(22.3) \quad ||pLOTS|| = 2 + 4 \frac{pLOTS}{15}$$

$$(22.4) \quad ||pHOTS|| = pHOTS.$$

In **the strict scoring approach**, the points are initially plotted in the interval [0, 6]. Scores below 2.5 are rounded to 2. Accordingly, membership functions for the second-level components are:

$$\begin{aligned} & \text{ELSE IF } (\|pHOTS\| - \|pLOTS\|) \\ & \quad \geq 2 \text{ THEN practice IS } \left(\frac{\|pHOTS\| + \|pLOTS\|}{2} - 0.5 \right) \\ & \text{ELSE practice IS } \left(\frac{\|pHOTS\| + \|pLOTS\|}{2} \right) \end{aligned}$$

(24.2) **theory**($\|tLOTS\|, \|tHOTS\|$):

$$\begin{aligned} & \text{IF } \frac{\|tHOTS\| + \|tLOTS\|}{2} < 2.5 \text{ THEN theory IS } 2 \\ & \text{ELSE IF } (\|tHOTS\| - \|tLOTS\|) \\ & \quad \geq 2 \text{ THEN theory IS } \left(\frac{\|tHOTS\| + \|tLOTS\|}{2} + 0.3 \right) \\ & \text{ELSE theory IS } \left(\frac{\|tHOTS\| + \|tLOTS\|}{2} \right) \end{aligned}$$

To determine the **final grade** in the presented **Variant 1**, it is applied that theory and practice are equal and the rules are formulated as:

(25) **finalR**(**theory**, **practice**):

$$\begin{aligned} & \text{IF } \frac{\text{theory} + \text{practice}}{2} < 2.5 \text{ THEN finalR IS } 2 \\ & \text{ELSE IF theory} = 2 \text{ OR practice} = 2 \text{ THEN finalR IS } 3 \\ & \text{ELSE finalR IS } \left(\frac{\text{theory} + \text{practice}}{2} \right) \end{aligned}$$

The final grade **finalN** is determined by rounding **finalR** to the nearest whole number:

(26) **finalN**(**finalR**) = **ROUND**(**finalR**)

Determination of HOTS ratings and LOTS in **Option 2 for HOTS - LOTS hierarchical multicomponent assessments** are also subject to similar reasoning and can be customized depending on the views of the particular assessor.

The formulas for evaluating the **practice components** and **theory** (24.1) and (24.2) are consistent with **Model 1**, therefore only components from the previous level are involved. In certain borderline cases, it is appropriate to use higher-level subcomponents and use **Model 2**.

An advantage of fuzzy logic in scoring is the ability to achieve accurate automated scoring according to the subjective logic of the evaluator. A major drawback is the difficulty for non-specialists to formalize the rules for forming the assessment.

CONCLUSION: the use of fuzzy logic guarantees obtaining correct estimates, with correctly written rules, while the use of artificial intelligence methods requires a sufficiently large number of suitable input-output samples for training. It should be noted that all automated approaches require teacher monitoring and control of the formed assessments. The proposed hierarchical multi-component assessment models aim at comprehensive assessment of various high- and low-order thinking skills, theoretical knowledge and practical skills, etc. They allow the use of fuzzy logic and can be adapted for use in assessing different disciplines, in different age groups.

A series of models related to monitoring and analysis of learners' activities in an educational environment have been created, presented in the publications [60, 70, 72, 97]. The generalized models for working with big data in a virtual laboratory space were created in [62, 63, 71].

A model through Generalized Net (GN) of the processes of personalization and usage of a learning environment is presented here (Figure 24). In this model, the students (learning objects) are interpreted by σ -tokens, the trainers are interpreted by τ -tokens, the serving administrative staff (study department inspectors) by α -tokens, the training courses (SCORM-packages) in the e-library by λ -tokens, the electronic services provided by the e-learning environment through γ -tokens and software agents, in the role of personal assistants to learners through β -tokens.

This model is a minimal reduced generalized network model [8, 9]. The dissertation gives a detailed description of the individual transitions.

The generalized net model consists of 11 transitions:

- Z_1 – entry of a new student (learner) into the electronic learning environment;
- Z_2 – entry of a training teacher into the electronic environment;
- Z_3 – training service from an administrative training department;
- Z_4 – introduction of new training courses in the electronic library;
- Z_5 – work with electronic services offered by the educational environment;
- Z_6 – selection of a training course by a student in the e-Learning environment;
- Z_7 – new training course administration process;
- Z_8 – work process of the agent (personal assistant of the student);
- Z_9 – student training process for the selected electronic course;
- Z_{10} – process of evaluating the student's education in the chosen course;
- Z_{11} – process of building educational content of an electronic course.

CONCLUSION: The created GN model [68] enables the tracking of personalization processes and the use of various intelligent tools for e-learning. Information about learner preferences and learning outcomes can be extracted. Based on the model and accumulated statistics from real data, assessments can be made and trends can be detected for the development of processes related to e-learning and its service. Additional model parameters and additional characteristics of the tokens can be introduced, taking into account the factors affecting the learning process, in order to increase the interest of the learners. The active participation of learners in the process of acquiring knowledge and acquiring skills can be greatly influenced by the quality of the educational space used.

4.2. A model of the process of data mining tools application in an e-Learning environments via Generalized nets

The created generalized net model describes possibilities for choosing and applying appropriate techniques for extracting knowledge from data in learning environments [59].

This model is a reduced generalized network model, but of a higher order than the model presented in Fig. 24, where an activation condition is defined for each transition. A boolean expression is set, if the value is "true" the corresponding transition can be activated, if it is "false" - no. The model presented in Fig. 25, contains 5 transitions and 21 positions, grouped into two groups and associated with two types of tokens that enter the corresponding types of positions: α -tokens and l -positions represent the process of extracting knowledge from the data, β -tokens and t -positions represent the criteria for limiting means and choosing appropriate techniques for extracting knowledge from data. For brevity, the notation α - and β -tokens instead of α_i - and β_j -tokens, where i, j are the numbers of the respective tokens.

Initially, an β_0 -token stands at position t_6 with an initial characteristic:
"means available to extract knowledge from data".

At the next transition from the functioning of the net, the β -token splits into two tokens. The original β -token will continue to stay at position t_6 , while the other β -token will move to the transition Z_5 , passing through the transition Z_3 .

The kernels α_0 and α_1 , entering to the net through positions l_0 and l_1 , receive characteristics,
 respectively:
"initial hypotheses";
"initial data".

Tokens β_1 and β_2 enter to the net through positions t_0 and t_1 . These tokens receive initial characteristics respectively:

"a new technique for extracting knowledge from data";
"criteria for selecting a data mining technique".

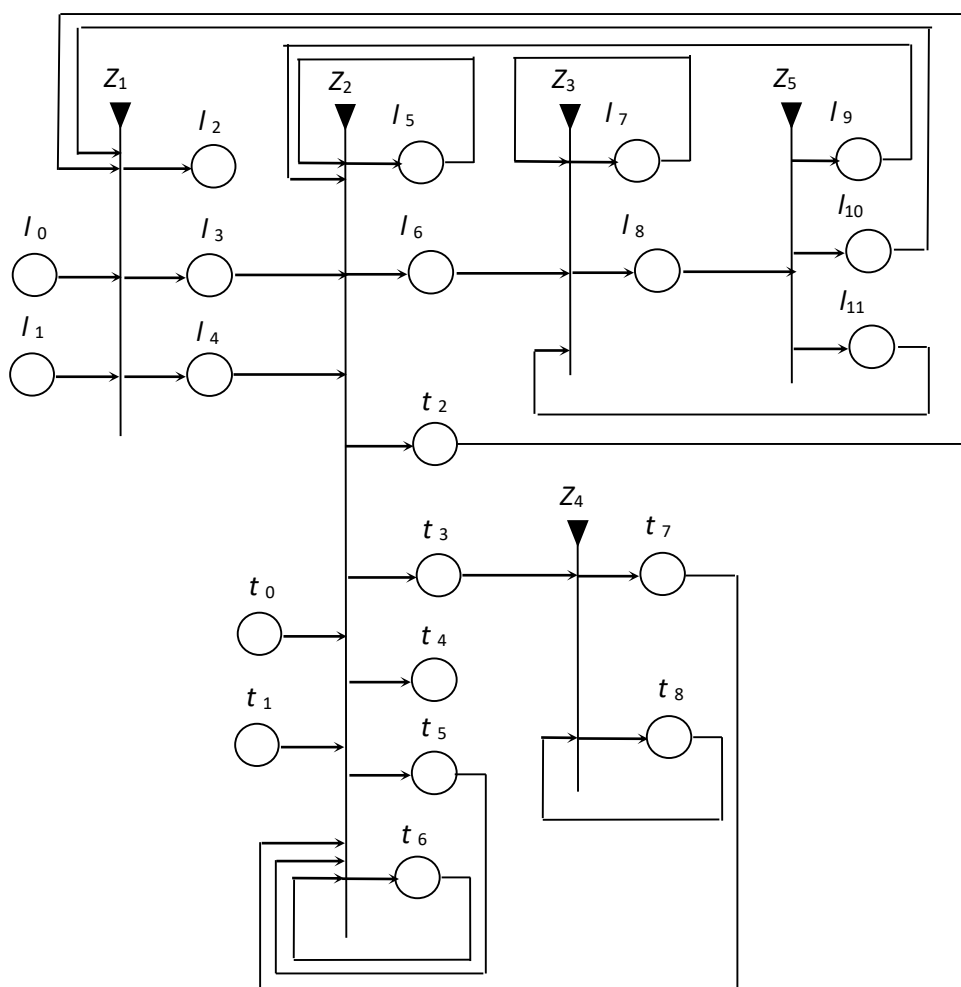


Figure 25. GN model of the process of data mining tools application in an e-Learning environments

The dissertation gives a detailed description of the transitions.

CONCLUSION: Integrating learning systems with data mining tools is necessary for the process of customizing e-learning and distance learning courses. Based on the obtained results, additional measures can be introduced to analyze and change the training courses and the analysis criteria. This, in turn, is a way to increase the quality of higher education.

4.3. A generalized net model of a multicomponent evaluation process

The objective assessment of students' knowledge and skills has many characteristics: true score, reliability *and* validity, honesty, differentiation, comprehensiveness of the assessment, etc. [91]. It provides educators with a better opportunity to both assess learner learning and analyze teaching effectiveness. A generalized net *model of the multi-component assessment process* was created, which includes six stages:

1. **Creating a meta-model of a test.** The meta-model defines types of assessment components defining what knowledge and skills are assessed according to a specific taxonomy and assessment models.
2. **Create a test pattern.** The test model includes characteristics for the target group, the subject of study, the start and duration of the test, etc. On each type of evaluation component of the meta-model is assigned a corresponding number of test units.
3. **Test Setup.** Specific test items are asked, which are entered or selected from a database of questions. For each specific test unit, it is indicated to which type of assessment elements it belongs. The specific trainees also ask.
4. **Conducting the test** is a process where specific tests are generated. They include the model-specified number of test units of each type. Learners enter their answers and solutions.
5. **Evaluation process.** It includes automatic scoring of the test according to the selected model and manual scoring of the open questions and tasks, if any.
6. **Analysis of the results and evaluation of the test.** The test creator uses the results to evaluate the test and the effectiveness of their scoring approaches.

The GN model describing the process of modeling and constructing tests for multi-component assessment of the knowledge and skills of learners [43] is presented in Fig. 26. This model is a minimal reduced generalized network model and contains the set of transitions:

$$A = \{ Z_1, Z_2, Z_3, Z_4, Z_5, Z_6 \}, \text{ where:}$$

- Z_1 – creating a meta-model of a test;
- Z_2 – creating a test model;
- Z_3 – test configuration;
- Z_4 – conducting testing;
- Z_5 – assessment process;
- Z_6 – analysis of the results and evaluation of the test.

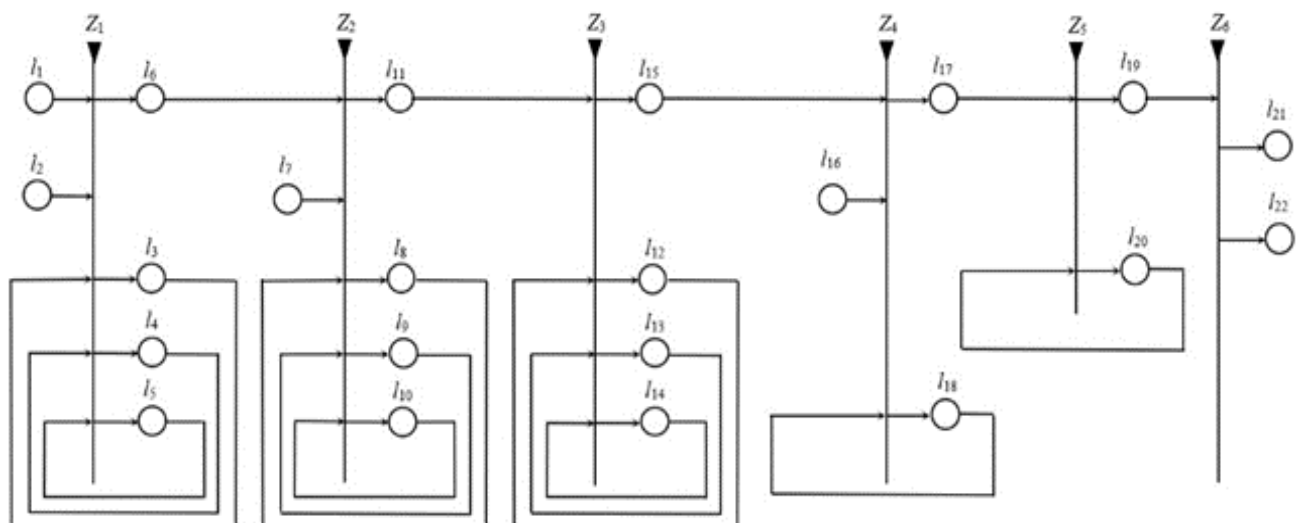


Figure 26. GN model of a multicomponent evaluation process

Tokens are used to describe the processes in the model:

- α - tokens interpret users - administrators, test creators, question authors, learners;
- β - tokens – types of assessment components, to assess practical or theoretical skills, low- or high-order thinking skills, skills categorized according to a specific learning taxonomy, etc.;
- γ - tokens interpret evaluation models – weighted average evaluation of components, evaluation based on weight coefficients, fuzzy logic, etc.;
- δ - token – test meta-model;
- η - token – test model;
- λ - token – data bank with test questions;
- μ - token– evaluation component (for example: test question, assignment, case study, etc.).

The process is started by a system administrator, represented by α - token, which enters to the net through position l_1 with initial characteristic:

"unique identifier and user name of the software system".

In position l_2 , a request is launched to create a new test meta-model, using a δ - token with an initial characteristic:

"unique identifier and name of test meta-model".

In position l_3 cycles α - token with characteristic:

"list of available software system users".

In position l_4 and in position l_5 cycle β - token and γ - token, respectively, with characteristics:

"types of assessment components" and

"available assessment models".

The dissertation gives a detailed description of the transitions.

CONCLUSION: A generalized net model is proposed, which aims, on the one hand, to generalize the assessment process, and on the other hand, to enable customization of the way of forming tests and assessment during training [43]. This has been achieved by defining meta-models and assessment models that set frameworks for creating specific tests and approaches to their assessment. The development of a software system implementing the described process will provide educators with a flexible platform for experimenting with various standard and proprietary approaches to test construction and student assessment.

4.4. A generalized network model of the processes in project-based learning

In its essence, project-based learning is a pedagogical model of interdisciplinary activities aimed at real problems [11]. Basic skills that the trainees form are: to learn to identify the stages in the development of a project, to plan their activities, to respect the planned deadlines, to work together with other team members, to evaluate the activities of other team members, to self-evaluate the actions themselves, to participate in the discussions on the topic of the project, forming and defending their own ideas and skills with arguments.

Project work management process in education

Successful management depends on the head of the discipline: he can allocate tasks for implementation, create a schedule for work, control the execution of subtasks and monitor the overall work. For this purpose, meetings can be held to present the current results and discuss the activities or work in a suitable interactive environment such as: Moodle, Acolad or an environment based on Wiki technology.

(1) The main activities of project definition

The main activities of project preparation and planning are formalized as:

- *Task specification for a specific project.*
- *Defining the sub-tasks, the time, for execution and who will deal with them.*
- *Defining and allocating resources.*
- *Managing the execution of tasks.*
- *Collection of various data for statistics and to measure the development of the project.*
- *A description of the possible risks for the project and a plan for its management.*
- *Generating various reports for the work on the project.*

(2) Planning teams (groups)

Main activities related to the planning of the groups working on the projects are:

- *Determining the number of participants in a team for a given project.*
- *Defining the structure of the team.*
- *Designation of a leader (designated by the teacher or chosen by the participants)*

(3) Setting the task before the students

When defining the project, the following must be presented to the students :

- *preliminary information;*
- *the task set in an interesting and motivating way;*
- *a step-by-step description of the process when performing the task;*
- *a set of recommended information sources;*
- *guidelines for organizing and storing information;*
- *ways and criteria for evaluating the performance of the task.*

(4) Activities of trainees performing projects

During the implementation of a project, the trainees have to: specify the question or task (for example, specify or summarize it); collect and analyze data from various sources; share, generate and discuss different ideas; make their own reasoned assumptions, hypotheses and predictions; conduct and analyze their own experiments; create artifacts (abstracts, databases, multimedia, models, prototypes); make proofs, generalizations and inferences; communicate and present their ideas and discoveries to other people; raise new issues for consideration and problems.

(5) Inspection and evaluation

Checking and evaluating the results of project work includes:

- *determining what exactly will be checked;*
- *determining by what means the inspection will be carried out;*
- *correctly cited sources;*
- *the public appearance as a natural conclusion for the developed project.*

If the teacher uses a data repository in which information about past projects has been collected, then after completion of work on current projects, the database must be updated. The accumulated information can be used to gather statistics about the students' skills and the study group as a whole.

Modeling the learning process through project work

In the generalized net model presented in Fig. 27, tasks (projects) are interpreted by α -tokens, teachers are interpreted by β -tokens, learners (learning objects) by γ -tokens, evaluation criteria respectively by δ -tokens [65]. This model is a minimal reduced generalized net model.

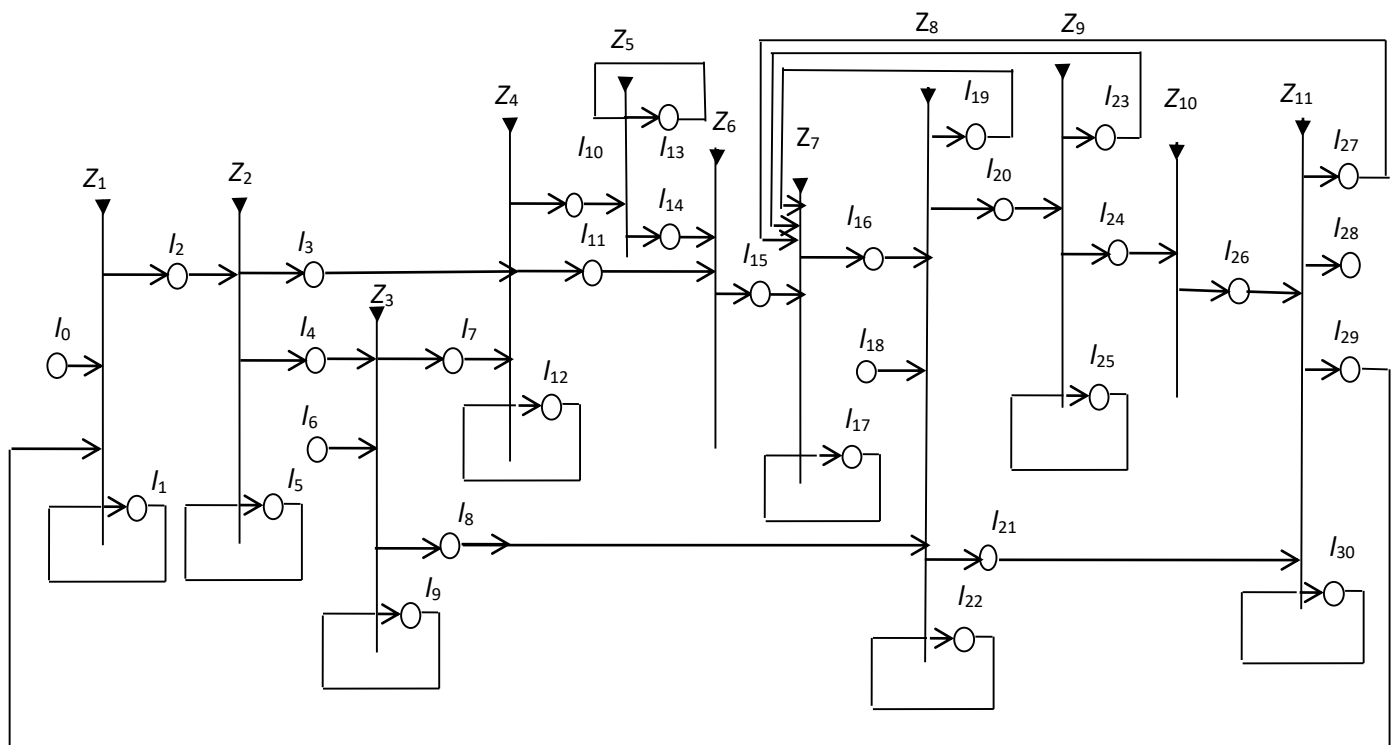


Figure 27. GN model of the processes in project-based learning

The GN model consists of 11 transitions with the following meaning:

- Z_1 = Preparation for project-based learning
- Z_2 = Start concrete project-based learning
- Z_3 = Forming teams of students
- Z_4 = Defining the tasks for each team
- Z_5 = Define the sub-tasks of each project
- Z_6 = Distribution of subtasks by team members
- Z_7 = Work on a specific subtask of the project
- Z_8 = Work process on the overall project
- Z_9 = Completion of project work
- Z_{10} = Presentation and protection of the project
- Z_{11} = Evaluation of projects

The dissertation gives a detailed description of the individual transitions.

CONCLUSION: The characteristics and possibilities of project-based learning in addition to traditional pedagogical approaches have been examined. A generalized net model of the learning process is built by developing a project [65]. The model serves to analyze the possibilities of working on projects adapted to the knowledge and skills of the students, the development of which leads to the integration of knowledge acquired in different courses of the study. The integration of project-based learning with the possibilities of electronic and web-based learning make the learning process more attractive and more diverse for students.

4.5. A model for the gamification of an e-learning course

Gamification of learning aims to integrate game elements and techniques into the e-learning process to support learning content. These games have educational, training or informational purposes. Gamification proposes that the learning process be organized into game levels (locked or unlocked with entry requirements) that present different sections with learning resources and activities to go through. The training course is introduced with an interesting context (plot/story) of the activities to be performed. Learning rules are introduced as game rules. Each level includes challenges - learning activities that must be completed to achieve the level's learning objectives. Some of the assignments are individual and others are collaborative, where learners work in teams.

For completing some activities, learners receive expected bonuses (e.g. points). For achieving a set of requirements, trainees can receive badges: Champion, Superstar, Adventurer, Explorer, etc. For excellent results or for performing a specific activity, they can receive unexpected rewards (additional interesting information, points/rating, material rewards (virtual objects) etc.). For a specific completed activity, benefits can be obtained as a combo (help, recommendations, more detailed examples, double points from an activity, etc.). Some of the learning items may be locked (hidden treasures) and unlocked when learners meet certain requirements.

Learners can participate in the gamified course with a specific game role, which also has a visual representation through an image (avatar). Based on the collected points and the current playing level, the learners are ranked in the ranking, where they can be the leading participants in the learning process. Throughout the process, learners have access to information about their learning progress as game progress and at any current moment about their status (game level reached, points earned, badges and other rewards).

A number of models have been developed to track the learning process and student outcomes [74, 78, 99]. The use of artificial intelligence methods enables prediction of results and partial automation of the evaluation process. A generalized network model for the processes of selecting and building a suitable E-Test system is given in [59]. A new approach for forming current assessments of learners' knowledge and skills based on fuzzy sets is presented in [39]. The current model generalizes these models and extends them with opportunities to implement modules for gamification of learning and to track learners as they work through a gamified course.

The created generalized model of the process of gamification of a training course is a minimal reduced generalized net model [69] and contains the set of transitions:

$$A = \{ Z_1, Z_2, Z_3, Z_4, Z_5, Z_6 \},$$

where:

- Z_1 – Creation of a model and gamification of training;
- Z_2 – Defining rules for learning in a gamified course;
- Z_3 – Conducting training with the electronic course;
- Z_4 – Appraisal of trainees and awarding;
- Z_5 – Ranking and analysis of student results;
- Z_6 – Learning analysis and course evaluation.

The following tokens are used to describe the processes:

- α - token - users of the learning environment (administrators, teachers, students);
- β - token – e-learning courses and gamification tools (plugins, software components, etc.);
- γ - token – evaluation models;
- η - token – data on the training objects during the training;
- k – token – specific data for the gamified modules for the learning objects;
- μ - token - rules and criteria for applying game elements and awarding.

The generalized net model of the process of gamification of a training course is presented in Fig. 28. The dissertation gives a detailed description of the individual transitions.

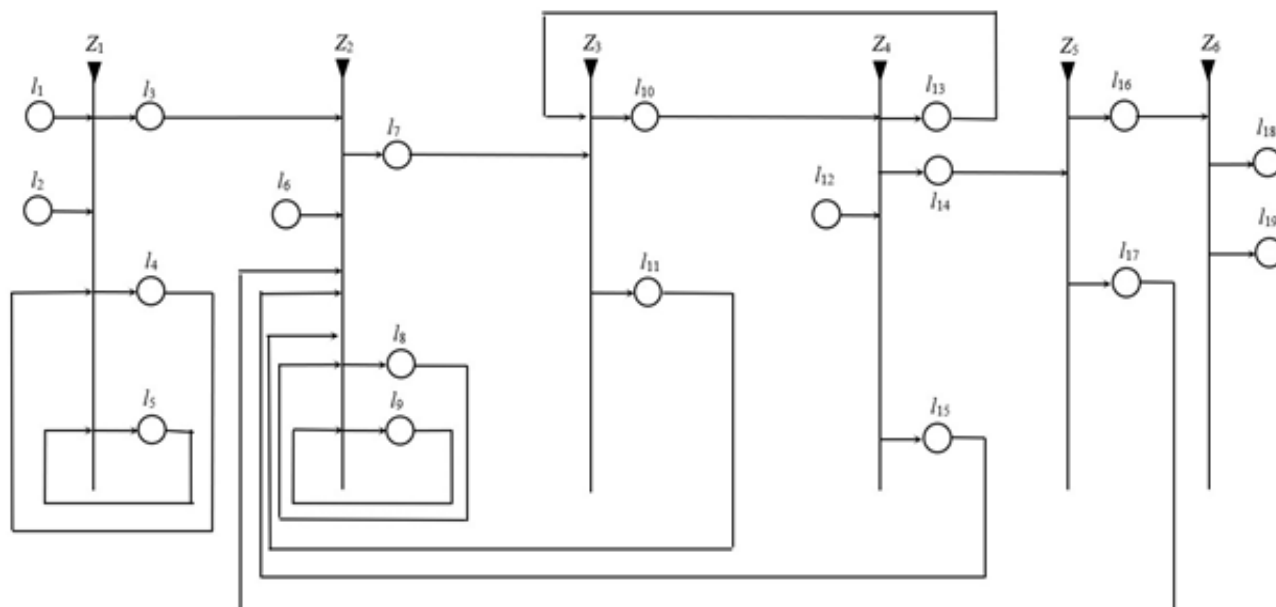


Figure 28 . A generalized net model of the gamification process of an e-learning course

The gamification of training is divided into two main types: structural change of a training course and content change of a training course.

With *the structural change*, the learners go through the standard learning resources, but various game elements are included in the course, such as: points, badges, leaderboards, avatars, medals and awards, etc.

In case of *substantive change* of a training course, game techniques are used to present the training content in the form of rules, levels and skills. For example:

- The rules of the learning process can be considered as the rules of a game;
- Mission/Challenge/Adventure – can be all the learning activities that the learner has to carry out within the learning course and to them can be added a game plot that describes the purpose of the mission;
- Hidden treasure - hidden learning resources that can be found/opened only when certain conditions are met (for example, when completing a mission);
- Plot / Story – the learning process, can describe a story with missions to complete.

Games have a strong motivational mechanism to achieve more effective learning. Examples of such environments are: GENIE, The Knowledge Arcade, TalentLMS, Frog, Expertus One, Accord LMS, Axonify, etc. [6, 44]. These systems are not freely available, but are offered as paid software. The Moodle environment provides opportunities to create a personalized learning environment and supplies a set of open source support tools, such as gamification modules can be used: LevelUp, Ranking block and Stash.

The created generalized network model of the process of gamification of a training course offers a solution to one of the problems formulated in the book of Academician Krasimir Atanasov "Generalized Nets and Intuitionistic Fuzziness in Data Mining" [34].

4.6. Models for quality assurance and accreditation of higher education

Quality assurance in higher education is a continuous process that requires a lot of effort on the part of educational institutions and accrediting organizations. There are two main forms: *internal and external quality assurance*. Higher education institutions build *internal quality assurance systems* by establishing their own rules and criteria for evaluating the educational process. They are also subject to *external evaluation - accreditation, which is carried out by licensed national and international agencies*.

The need for commensurability of the quality of higher education in the European Union motivates the definition of uniform quality standards. In 2005, Standards and Guidelines for Quality Assurance in the European Higher Education Area (ENQA) were developed. The document (ESG) [18] contains numerous standards and guidelines for their implementation. Bulgaria is one of the countries that apply the ESG standard. Quality assurance of higher education in Bulgaria is carried out by the National Assessment and Accreditation Agency (NAAA) [2]. *Quality assurance procedures* follow a general scheme and include basic activities: self-assessment, expert group visit, assessment and post-accreditation control.

Assessment and accreditation procedures of higher education should be a natural and integral part of the educational process and should be carried out without much effort, imperceptibly for students and teachers. For this purpose, it is necessary for educational institutions to maintain repositories with up-to-date information about their activities, and accreditors - to provide software tools for automated quality assessment, agreed with the information in the national center for information and documentation (NCID). Models of the current state have been created in [40, 43]. A general model of the current state is presented in Fig. 29.

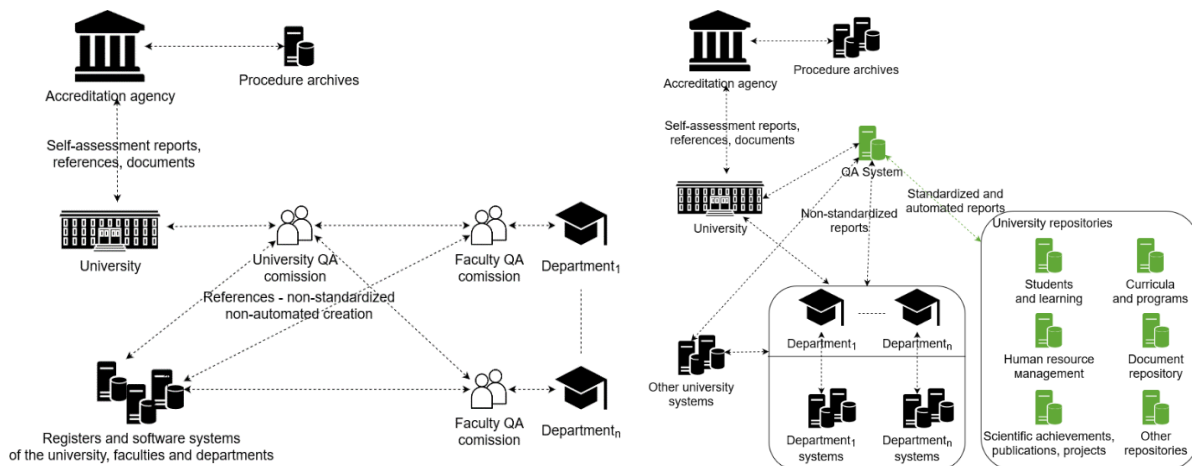


Figure 29. Model of the processes for preparing a self-assessment report by higher schools

- a) Model without university software QA quality system
- b) Model with a university software QA quality system

Many higher education institutions lack a system supporting the creation of a self-assessment report (Fig. 29 a). The allocation of tasks for its preparation becomes hierarchical: NAOA opens an accreditation procedure and informs the university; the accreditation commission allocates tasks for drawing up references and preparing texts for individual standards; units collect information from various registers, university software systems and also non-digitized information. The creation of the self-assessment report is difficult due to the need to summarize the data provided in different formats and bring the texts into a common document. The availability of tools supporting the activity (Fig. 29 b) provides opportunities for: control and tracking of activities; automation of access to references, etc.

4.6.1. A weakly-centralized centralized accreditation model

One possible approach to automation of accreditation procedures is by *building a centralized system of the accreditation agency and independent university quality assurance systems* (Fig. 30). Each university QA system automatically retrieves the required credential data from university repositories [45]. Another part of the information is entered by university officials. At a certain point, the information from the university's QA system is exported to the QA system of the accrediting institution.

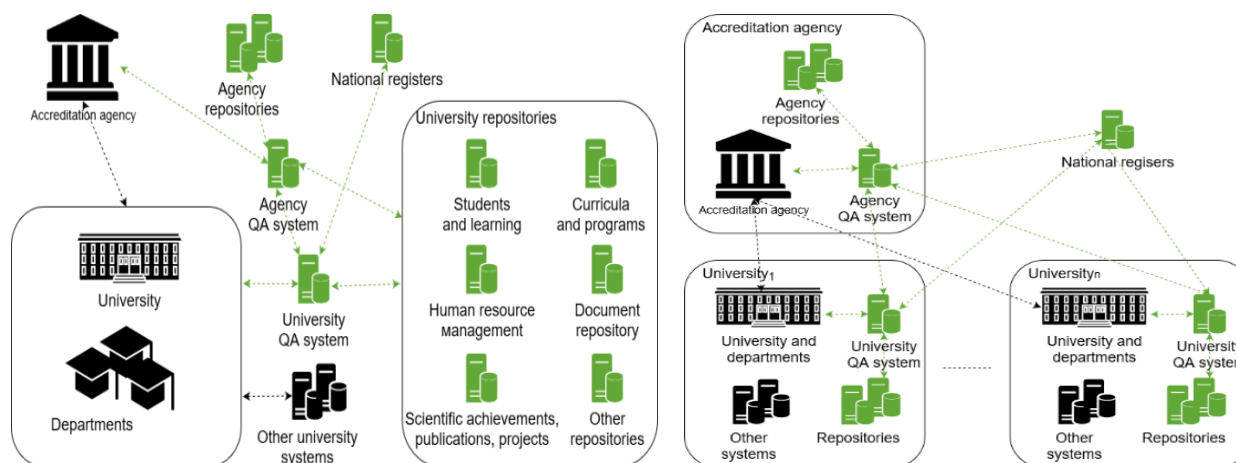


Fig. 30. A loosely centralized accreditation model

a) Main systems and communication interfaces in the accreditation process of a higher education institution

b) Main systems and communication interfaces in the accreditation process in the higher education system

The advantages of this approach also determine its disadvantages. The creation of a proprietary solution or the maintenance of a standardized application by universities requires the presence of specialized IT specialists in each higher education institution. Establishing a standardized university QA system is a labor-intensive task that must accommodate the presence of many different university repositories and other software systems.

4.6.2. A highly centralized accreditation model

The highly centralized model offers *the use of a centralized cloud system for accreditation*. The system supports a module for both the agency and configurable modules serving the accreditation processes for individual higher education institutions. The use of a single system implies simplified integration between the main modules, user-friendly software update, possibilities for integration of future software extensions, creation of dynamic analyzes and references on the data provided by all institutions [40].

The main purpose of the described models is to offer a *platform* that facilitates the work of all participants in the assessment and accreditation procedures - on the part of the trainers and on the part of the accrediting institutions. The use of software systems based on the proposed models has many advantages: centralization of evaluation processes, which facilitates their administration and monitoring; providing a common platform for team asynchronous work of users with different rights and responsibilities; possibility to work at any time and from any place; facilitated communication and transparency; automated documentation generation; optimization of time and material resources; reduce the probability of making mistakes, etc.

4.6.3. A generalized net model of data processing in self-assessment in higher education

A model of a centralized system for quality assurance of higher education was created in [42]. A generalized net model of a quality assurance system in a higher education institution is presented in [41]. Here, a minimal reduced generalized net model describing the data processing of a self-assessment procedure in a higher education institution is created, which contains the following set of transitions:

$$A = \{Z_1, Z_2, Z_3, Z_4\},$$

where:

- Z_1 – Collection of data from various sources;
- Z_2 – Data Integration;
- Z_3 – Data processing for the requirements related to the criteria system;
- Z_4 – Preparation of self-assessment report and relevant attachments.

The following tokens are used to describe the processes:

- α -tokens – data on the activities of the higher school;
- β -tokens – members of the administrative and academic staff, members of the quality committee and the group for preparing the self-assessment report;
- μ -tokens – criteria and criteria system of the accrediting institution;
- η -tokens – a data repository with up-to-date information on the activities of the higher school, necessary for the preparation of a self-assessment report;
- γ -tokens – data processing tools;
- σ -tokens – data requests from user/application.

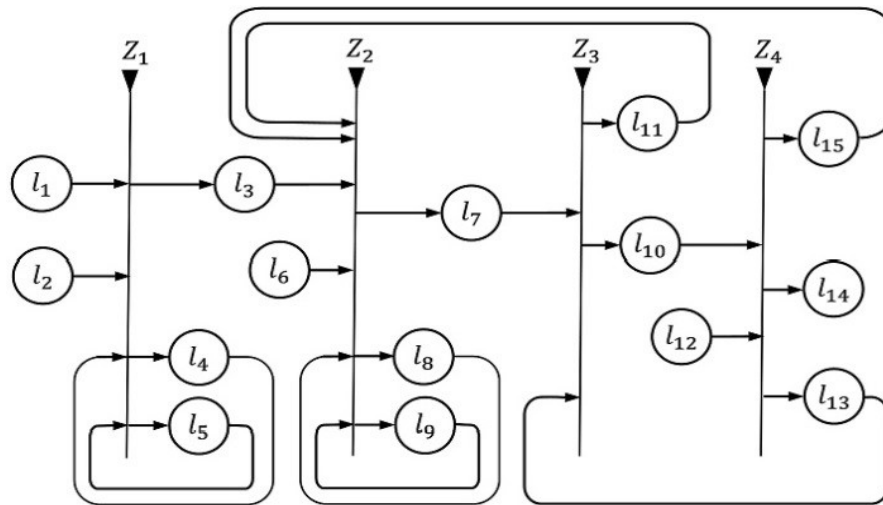


Figure 31. GN model of data processing in self-assessment in higher education institutions

The generalized net model describing the data processing in the self-assessment procedure in the higher education institution is presented in Fig. 31. The dissertation gives a detailed description of the individual transitions.

CONCLUSION: A formal model is proposed describing the means of data collection and processing of a self-assessment procedure in a higher education institution. This model can also be developed by using a **hierarchical operator** (H3 from GN theory) that replaces a given transition or position with a sub-net that has the same but described more detailed element behavior. Based on the created generalized net model and collected data from real processes, behavioral patterns and performance analysis of individual components can be discovered.

CHAPTER 5 . SOLUTIONS OF SCIENTIFIC APPLIED PROBLEMS IN EDUCATION

Results of applications of the research and analysis in an educational environment are presented in several author publications. Methodical techniques for introducing basic concepts in courses for students, adapted to the knowledge and specialties of the learners, are proposed in [64, 81]. How to follow current trends in data analysis courses by solving appropriately selected real-world problems and tasks is discussed in [67]. An approach to presenting concepts and standards for information security in higher education is proposed in [73]. Programming with reactive blocks is presented in [66] for Internet of Things courses. In [98], the difficulties in learning by deductive methods for program verification and synthesis are analyzed. Trends in the education of software engineers are presented in [34, 35]. This thesis chapter summarizes findings from these publications and proposes key curriculum elements and approaches for introducing data science core modules and tools in Higher Education.

5.1. Data Analysis curriculum

In determining the content of the course taught, each major and specific major should focus on the goals they set for their student education. On the one hand, applied majors emphasize the ability of students to be able to choose appropriate software for data analysis. In other majors, the emphasis is on learners to master basic algorithms for data analysis, machine learning and artificial intelligence. The student should be able to construct models to analyze an existing situation and future forecast, learn how to use various artificial intelligence techniques to detect anomalies and create optimal models. In the most ambitious plan, **after completing the training the student must be able to manage the entire data cycle and extract knowledge from it using methods of data collection and analysis and use of appropriate algorithms and data management systems.**

Taking into account the specifics of training in the direction and specialty, the curriculum of the discipline related to data analysis can emphasize different aspects. The following is a proposal of **core modules** that may find a different place in training in the discipline of data analysis (regardless of the specific name).

(1) Preliminary Preparation. Students should have taken the core courses in programming, mathematics, probability theory and statistical methods, databases, artificial intelligence, discrete mathematics, data structures and communication to enable the learner to relate the methodologies to the introduced new knowledge of the discipline.

(2) Introductory. This **module** should introduce the basic stages in the data mining and analysis cycle and review various tools that support each stage of the cycle. The aim of the module is to help students in the future to successfully realize themselves as data analysts or to conduct scientific research and develop an academic career.

(3) Main part on the course. This **module** should provide the theoretical training with the basic algorithms and approaches for data analysis such as: classification analysis, regression analysis, associative analysis, cluster analysis, noise analysis, etc. A main emphasis in the process is the application of a relevant Data Mining algorithm or algorithms, enabling the acquisition of knowledge describing: relationship between data properties, data models, the results of data classification and clustering, etc. It is interesting what types of dependencies can be discovered in the process of extracting patterns from data. For this purpose, six types of tasks are considered, the solution of which leads to obtaining the dependencies we are interested in:

- The task of *describing and summarizing data*.
- *The segmentation (or clustering) task*.
- *Analysis of extremes (outlier detection)*.
- *Concept description task*
- *Classification and regression*.
- *Dependency analysis*.

Different mathematical tools and artificial intelligence methods are used to extract and represent the data patterns. The main methods used by the techniques for extracting regularities from data are: method of Support Vector Machines (SVM), method of Naïve Bayes, decomposition by the method of Principal Component Analysis, method of Artificial Neural Networks, etc.

(4) Hands-on learning – learners experiment with data mining, data management and analysis, data visualization approaches, by implementing various open source systems. The course should introduce popular systems for data processing and data mining, such as: *WEKA* (Waikato Environment for Knowledge Analysis) [103], *RapidMiner* [87], *KNIME* [52], *Orange* [56] and others.

(5) Course project. Students with weaker programming skills and insufficient mathematical knowledge will experience a number of difficulties in understanding the essence of data analysis and the application of the chosen technique. To overcome the difficulties, the inclusion of project-based learning is proposed as the last module of the course. Course projects can be individual or team, but with a specific orientation to a certain business area. For example:

- organizational and business data analysis: financial data, insurance, banking, investment management, risk management, market data, etc.;
- the analysis of public health data: data on the quality of life of the person, construction of models for quality of life, prevention of diseases and formulation of community health management programs;
- the management of public institutions, hospitals, state institutions, etc.;
- data analysis in our smart cities and cyber-security.

Whether the goal is to discover interesting relationships, categorize objects into groups, optimize resource planning, or determine billing rates, project work will help students gain a basic understanding of data analysis techniques and acquire knowledge extraction skills with specific tasks.

CONCLUSION: The development of such academic programs is critical to the success of learners as future researchers in the digital world. The interest of learners and their active participation in the learning process can be influenced by the quality of the learning process and the curriculum used.

5.2. Data mining tools through examples

Database courses for the students of *Informatics and Computer Sciences in most Bulgarian universities takes place* in the second or third semester. For this purpose, the following are successively introduced: basic data models; design process of relational databases, analysis and normalization of relational schemas, languages for data description and processing. Students become familiar with the problems of distributed databases and object-oriented databases and the new perspectives for the development of big data (BigData). After implementing a number of applications in the form of course projects, students move on to analysis of data warehouses and Data mining techniques. These techniques offer rich opportunities for data exploration, but learning them requires a strong mathematical foundation. Considering the limited time, the courses here offer to reveal the possibilities of these techniques by offering students specific examples on previously prepared data. Students can analyze data following the methodology of the CRISP-DM model, following the six main phases in the process of extracting patterns from data in a six-step sequence. For this purpose, appropriately selected real tasks for data analysis are used, which have been adapted for the purpose of easier application of a corresponding algorithm, presented in detail in the dissertation work.

CONCLUSION: The analysis of the results of the conducted training showed an increase in the degree of assimilation of the educational content and an increase in interest in the discipline [67]. The study of topics did not cause serious difficulties among students who have the necessary preliminary training. Students with weaker programming skills and insufficient mathematical knowledge experience a number of difficulties in understanding the analysis and applying the chosen data mining tools.

5.3. Introducing ontologies to student learning

A method for introducing ontologies into the courses of majors in the field of "Informatics and Computer Sciences" is proposed through realization of three main tasks that are solved in the learning process [64]. The first task is related to describing the nature of the ontology. The origin and meaning of ontology and the different types of ontologies are presented. Students become familiar with the main components of the ontology structure. The second task is related to the representation of schemas that implement ontologies. This part of the tutorial is based on examples, a comparison between different example ontologies, as well as an analysis of each of them. To do this, the tools of the OWL language are introduced. The third task is the development of an example OWL ontology, using Protégé software.

The introduction of the concept of "ontology" begins with an introduction to its origins in philosophy and reflects the nature of things that actually exist. In modern philosophical literature, this term is used to **denote a certain system of categories that are a consequence of a certain system of views (a certain point of view) about the world**. Their classification is presented depending on different classification features. Some authors consider ontologies as concept-oriented, others as object-oriented in the subject area. In [38] a classification is considered, according to which ontologies are divided into: general ontologies; ontology oriented to a specific field; task-oriented ontologies and applied ontologies.

Regarding the structure of the ontology, two components are clearly distinguished: names of existing concepts and relations in the domain. A number of restrictions can be imposed on domains. The ontology together with many concrete instances of the classes constitutes a knowledge base. Thus, the development of an ontology goes through the following phases [36]: definition of the classes in the ontology; building a hierarchy of classes; defining characteristics and describing their possible values; populating specific values for the characteristics to obtain specific instances.

Ontology development is an iterative process. A basic rule is that the concepts in the ontology are close to the objects and relationships in the specific domain. When building an ontology, it must be determined in advance - for what purpose the ontology is being created, what types of questions are supposed to be answered (with its help), how it will be used and maintained. The training continues with the study of: the ontology design methodology; learning Web Ontology Language (OWL) and building a sample ontology; using Reasoner to work with ontologies.

For training purposes, a concrete example of an OWL ontology is developed with the Protégé software. Specific activities performed by students through Protégé are:

1. Create classes. After the classes are created for example: *Student* and *Teacher*, can be made *disjoint* - meaning that an object cannot be an instance of more than one of these two classes.
2. Create a class hierarchy. For example: the subclasses - *Lectures* and *Assistants*.
3. Defining OWL *Properties* that represent relationships: *Object properties* and *Data properties*.
4. The next step is to introduce OWL characteristics of *object properties* (*inverse property*, *functional properties*, *transitive*, *symmetric* and *antisymmetric*, *reflexive*). Property *Domains* and *Ranges* associate domain entities with range entities. Property *Restrictions* is a class of individuals that is described or defined by the relationships in which those individuals participate.
5. The next step in the learning process is the definition of *primitive* and *definite* classes, based on necessary and sufficient conditions. OWL allows all elements of an ontology, as well as the ontology itself, to be documented with metadata.

As a result, an introduction to the ontologies is achieved, by explaining the basic concepts and the main principles in the field, based on examples. The ontological system is developed together with the students, going through all the steps in the creation process. In the learning process, opportunities are provided to complement the built ontological system - with new concepts, relations or attributes.

CONCLUSION

Educational Data Mining and Learning Analytics are relatively new fields that aim to improve the educational experience, help stakeholders (teachers, students, administrators, researchers), take better decisions using accumulated data. Despite the great expectations and the increased volume of publications in the field of data mining in educational environments, there are still barriers and challenges facing researchers in the field, such as the lack **of comprehensive and easy-to-use and understand tools that can be integrated into the most popular learning management environments** .

Modern education uses software platforms for the management of educational content, which provide a technological apparatus, in parallel with the publication of educational content, to collect and store information about the activity of users - teachers and students. A direction for future development of the work is the creation of easy-to-use software tools, with the possibility of integration into learning management environments, for the early detection of learners at risk and prompt notification to teachers of which learners need additional help and which teaching practices greatest impact. In addition, research should continue on the possibilities of personalizing the learning process and content so that each learner receives resources according to his current knowledge and attitude towards the learning process.

A challenge is to develop adaptive courses that are automatically customized according to learner profiles (needs, goals, background, country, learning style, etc.). The focus of the dissertation is on monitoring and identifying at-risk learners who are likely to drop out or fail during training. But the data analyzes conducted are on relatively small groups of learners. Better results can be obtained by analyzing a large number of students, courses and institutions. For example, MOOCs (Massive Open Online Courses) can use data from thousands of students. Participants have different backgrounds, maturity, experience, education levels, language skills, goals, needs and learning styles. But a significant barrier [37] is related to **the ethics of using personal data, which must be taken into account at all stages - from data collection to interpretation of results and decision-making**. For example, data related to gender, social status, race, religious beliefs, ideology or disability could lead to discriminatory treatment. On the other hand, the series of measures related to the protection of personal data must also be taken into account.

The next step of research in the field is the creation and application of methodologies for the use of Big Data Analytics in educational environments. In the era of big data, the possibilities of storing, managing and processing data from online learning environments make it possible to better study learning processes and look for effective ways to improve them. **The combination of Big Data and Learning Analytics is a promising area of research.**

Dissertation Contributions

A. Scientific contributions (in the class enrichment of existing knowledge)

Methods and models are proposed, as a result of theoretical summaries of the processes of observation and analysis of the activities of learners:

1. A five-step method for assessing and predicting the knowledge, skills and competencies of learners in the virtual educational space, with the possibility of applying computational models and criteria for dynamic assessment, and the final assessment is formed by multiple assessment components. The conducted empirical studies confirm that the proposed method is promising for the development of an early warning system for various stakeholders of the learning process.
2. Fuzzy logic models of hierarchical organization of evaluation components and dependencies between them that the evaluator explicitly or implicitly uses in evaluation. Two models are available: *Model 1* of hierarchical tree-like organization of evaluation components and *Model 2* of hierarchical graph organization of evaluation components.
3. A method with Web metrics and Inductive Fuzzy Classification for evaluating the degree of use of websites by learners, when analyzing their behavior in the learning environment and the web space.
4. A method for document type analysis, based on classification algorithms, with the target attribute being whether the document is related to the field of study. Data analysis makes it possible to look for the difficulties that learners encounter when working with literary sources and to design modules to meet their individual needs when searching and using documents.
5. Three-factor model of the learner, which includes: competency factors, emotional factors, social environment impact factors. Through the learner model, the behavior and change of knowledge, skills and competencies can be tracked and predictions can be made, according to the accumulated data from the operation of e-learning environments and systems with different users.
6. A generalized net models: with the ability to track the processes of using different tools in e-learning environments; the process of applying Data Mining tools in Learning environments; a comprehensive multicomponent evaluation process in six stages with possibility to customize the way of forming tests by defining meta-models setting frameworks for creating tests and approaches for their assessment; the processes in project learning and the possibilities for integration with E-learning and web-based learning; E-learning gamification course to analyze the possibilities and problems in learning with game situations. The created models offer a partial solution to the problems formulated in the book of Acad. K. Atanasov „Generalized Nets and Intuitionistic Fuzziness in Data Mining“ [9].
7. Weakly-centralized and highly-centralized model for quality assurance and accreditation in higher education. Basic generalized net model of information flows in the processing of data in self-assessment in higher education.

B. Scientific applied contributions (in the class of application of scientific achievements in practice)

Proposed are:

1. Multi-step methods for creating models in an educational context using the machine learning tools of the *Orange Mining Data system*.
2. Software tools for analyzing and comparing the behavior of two machine learning agents: *Rule-Based System* and *Reinforcement Learning*.
3. Software tools for the analysis of sound frequencies and their conversion into colors in the RGB model. A model of a system that implements the conversion of sound frequencies into color.
4. Core modules of a Data science curriculum in Higher Education disciplines.
5. Methodical techniques for teaching students with data mining tools through examples from real problems and tasks for designing and implementing ontologies in student learning.

**List of scientific publications on the topic of the dissertation work
in publications that are referenced and indexed in world-famous databases with scientific information
(Web of Science and Scopus)**

- 1 . **Orozova D.**, Appropriate E-Test System Selection Model, Comptes rendus de l'Académie bulgare des Sciences, 2019, Vol 72, No 6, pp. 811-820, ISSN 1310-1331, DOI: 10.7546/CRABS.2019.06.14 , **IF** =0.343 (2019), **WoS Q4** , **SJR** =0.218, (Scopus Q2)
<https://www.webofscience.com/wos/woscc/full-record/WOS:000477784300014>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85070988936&partnerID=MN8TOARS>
2. **Orozova D.**, K. Atanasov, Generalized Net Model of Processes Related to Big Data, Comptes rendus de l'Académie bulgare des Sciences, book No 12, vol. 71, 2018, pp. 1679 -1686 ISSN 13101331, DOI: 10.7546/CRABS.2018.12.13 , **IF**= 0.321 (2018), **WoS Q4**, Scopus **SJR** =0.205, (Scopus Q2)
<https://www.webofscience.com/wos/woscc/full-record/WOS:000456750000013>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85063227040&partnerID=MN8TOARS>
3. **Orozova D.**, K. Atanasov, Model of Big Data Map/Reduce Processing, Comptes rendus de l'Académie bulgare des Sciences, 2019, Vol 72, No11, pp.1537-1545, ISSN 1310-1331, DOI: 10.7546/CRABS.2019.11.11, **IF** =0.343 (2019) , **WoS Q4**, **SJR** =0.218, (Scopus Q2)
<https://www.webofscience.com/wos/woscc/full-record/WOS:000502809100011>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85109014572&partnerID=MN8TOARS>
4. **D. Orozova**, K. Atanasov, M. Todorova, Generalized Net Model of the Process of Personalization and Usage of an e-Learning Environment, Proceedings of the Jangjeon Mathematical Society 19 (2016), No. 4, pp. 615 – 624. ISSN (print):1598-7264, ISSN (online):2508-7916, ISBN:89-87809-15-3, **SJR** = 0.508, (Scopus Q2)
<http://www.scopus.com/inward/record.url?eid=2-s2.0-84996565258&partnerID=MN8TOARS>
5. **Orozova D.** , M. Todorova, Ontology concept in courses on students, TEM Journal. Volume 7, Issue 3 , Pages 693-697, ISSN 2217-8309, DOI:10.18421/TEM73-29, 2018. **SJR** = 0.148. (Scopus Q4)
<https://www.webofscience.com/wos/woscc/full-record/WOS:000442699500029>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85052218825&partnerID=MN8TOARS>
6. Popchev I., **D. Orozova**, Towards Big Data Analytics in the E-learning Space. Cybernetics and Information Technologies, Vol. 19(3), 2019, pp.16-25, ISSN: 1311-9702, DOI: 10.2478/cait-2019-0023, **SJR** =0.31, (Scopus Q2)
<https://www.webofscience.com/wos/woscc/full-record/WOS:000486349500002>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85073873211&partnerID=MN8TOARS>
7. **Orozova D.**, K. Kaloyanova, M. Todorova, Introducing Information Security Concepts and Standards in Higher Education, TEM Journal. Volume 8, Issue 3, Pages 1017-1024, ISSN 2217-8309, DOI: 10.18421/TEM83-46, 2019, **SJR** =0.167, (Scopus Q2)
<https://www.webofscience.com/wos/woscc/full-record/WOS:000483978200046>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85072752036&partnerID=MN8TOARS>
8. Popchev I., **D. Orozova**, Towards a multistep method for assessment in e-learning of emerging technologies, Cybernetics and Information Technologies, Volume 20, No 3, Sofia, 2020, pp.116-129, ISSN: 1311-9702 , DOI: 10.2478/cait-2020-0032 , **SJR** = 0.272, (Scopus Q2)
<https://www.webofscience.com/wos/woscc/full-record/WOS:000572532500009>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85093080750&partnerID=MN8TOARS>

9. Hadzhikolev, E., S. Hadzhikoleva, K. Yotov, **D. Orozova**, Models for Multicomponent Fuzzy Evaluation, with a Focus on the Assessment of Higher-Order Thinking Skills, TEM Journal, Vol.9, No.4, pp .22-28, 2020, ISSN: 2217-8309 , DOI: 10.18421/TEM94-43, SJR =0.199, (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000596297300043>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85092629596&partnerID=MN8TOARS>

10. Todorov, J., V. Valkanov, S. Stoyanov, B. Daskalov, I. Popchev, **D. Orozova**, Chapter 6: Personal Assistants in a Virtual Education Space, Sgurev V., Jotsov V., Kacprzyk J. (eds) Practical Issues of Intelligent Innovations. Studies in Systems, Decision and Control, 140, Springer International Publishing AG part of Springer Nature, 2018, Library of Congress Control Number 2011893667, ISSN 2198-4182 , DOI: 10.1007/978-3-319-78437-3_6, SJR =0.131, (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000466552900007>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85049662711&partnerID=MN8TOARS>

11. Todorov J., E. Doychev, **D. Orozova**, A. Stoyanova-Doycheva, IoT Multiagent Assistant in Virtual Educational Space, 2018, AIP Conference Proceedings Vol. 2048, pp.020031; ISSN: 0094243X, ISBN: 978-073541774-8, doi:10.1063/1.5082049, <https://doi.org/10.1063/1.5082049>, SJR =0.182 (for 2018).

<https://www.webofscience.com/wos/woscc/full-record/WOS:000468108800031>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85058786246&partnerID=MN8TOARS>

12. Stoyanova -Doycheva A., T. Glushkova, S. Stoyanov, **D. Orozova**, Lifelong Learning Supported by an Intelligent Tourist Guide, 2018, AIP Conference Proceedings, Volume 2048, pp. 020038; ISSN: 0094243X, ISBN: 978-073541774-8, SJR =0.182 , DOI: 10.1063/1.5082056

<https://www.webofscience.com/wos/woscc/full-record/WOS:000468108800038>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85058819893&partnerID=MN8TOARS>

13. Hadzhikolev E., S. Hadzhikoleva, **D. Orozova**, K. Yotov, A Comprehensive Approach to Assessing Higher and Lower Order Thinking Skills, Uncertainty and imprecision in Decision Making and Decision Support: New Advances, Challenges, and Perspectives, Lecture Notes in Networks and Systems , Volume 338 , pp. 164–177, Lecture Notes in Networks and Systems, Volume 338, pp. 164–177, ISSN 2367-3370 , 2022, DOI: 10.1007/978-3-030-95929-6_13, SJR =0.151, (Scopus Q 4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000775291100013>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85126177886&partnerID=MN8TOARS>

14. Hadzhikoleva S., **D. Orozova**, N. Andonov, E. Hadzhikolev, Generalized Net Model of a System for Quality Assurance in Higher Education, 45th International Conference Applications of Mathematics in Engineering and Economics (AMEE'19), 2019, pp. 040005-1 -040005-8, ISBN: 978-0-7354-1919-3. AIP Conference Proceedings, 2019, DOI: 10.1063/1.5133515, Part of ISSN: 15517616 0094243 X , SJR =0.19.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000521744400033>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85075796767&partnerID=MN8TOARS>

15. Orozov B., **D. Orozova**, An AI application in E-Learning, CEUR Workshop Proceedings 3061, pp. 76-80, September 27–28, 2021, Plovdiv, Bulgaria, ISSN 1613-0073 , SJR =0.228 (for 2021) .

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85122281151&partnerID=MN8TOARS>

16. Ivan Popchev, **Daniela Orozova**, Algorithms for Machine Learning with Orange System, International Journal of Online and Biomedical Engineering (iJOE), Vol. 19, No. 04, 2023, pp. 109 -123, ISSN: 2626-8493, SJR =0.28, (Scopus Q3), DOI: <https://doi.org/10.3991/ijoe.v19i04.36897> .
<https://www.webofscience.com/wos/woscc/full-record/WOS:000967834800007>
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85152704594&origin=resultslist&sort=plf-f>
17. **D. Orozova**, N. Hristova, An application of analytical data research in e-learning system, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering Volume 1031 (2021) 012058, IOP Publishing, Online ISSN: 1757-899X, DOI: 10.1088/1757-899X/1031/1/ 012058.
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85101662489&partnerID=MN8TOARS>
18. Popchev I., **D. Orozova**, DataScience: Experience and Trends, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering, Volume 1031 (2021) 012057, IOP Publishing, Online ISSN : 1757-899X, Print ISSN: 1757-8981, doi:10.1088/1757-899X/1031/1/ 012057.
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85101718405&partnerID=MN8TOARS>
19. M. Todorova, **D. Orozova**, Training Difficulties in Deductive Methods of Verification and Synthesis of Program, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 9, No. 7, 2018, pp. 18-22, ISSN: 2158107X ISSN (online):2156-5570, doi:10.14569/IJACSA.2018.090703.
<https://www.webofscience.com/wos/woscc/full-record/WOS:000441330600003>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85054031652&partnerID=MN8TOARS>
20. **D. Orozova**, SQL User Queries Execution Model , Advances in Intelligent Systems and Computing book series (AISC, volume 1081), Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives, Volume 1081 AISC, 2021, Pages 184-189, ISSN 2194-5357 , DOI: 10.1007/978-3-030-47024-1_20
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85087754542&partnerID=MN8TOARS>
21. Orozov B., **D. Orozova**, Rule Based System Against Reinforcement Learning, CompSysTech'21: ACM International Conference Proceeding Series on Computer Systems and Technologies '21, June 2021, Pages 67–70, ISBN: 978-145038982-2 , DOI: 10.1145/3472410.3472437.
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85117607164&partnerID=MN8TOARS>
22. **Orozova D.**, M. Georgieva, A model of the process of Big Data with generalized net, 2016, IEEE 8th International Conference on Intelligent Systems, IS 2016 – Proceedings, 7 2016, Article number 7737487, pp.599-603, ISBN: 978-150901353-1, DOI: 10.1109/SIELA.2016.7543030
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85006035368&origin=resultslist&sort=plf-f>
23. Hadzhikolev E., S. Hadzhikoleva, **D. Orozova**, Digital Model of a Document in a University Document Repository, IEEE XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018, 3-6 June 2018, Bulgaria , ISBN 978-1-5386-3418-9 , DOI: 10.1109/SIELA.2018.8447089.
<https://www.webofscience.com/wos/woscc/full-record/WOS:000454837100037>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85053840788&partnerID=MN8TOARS>

24. **Orozova D.**, A. Ivanov, Generalized Net Model of Virtual Collaboration Space, IEEE XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018, 3-6 June 2018, Bourgas, Bulgaria, ISBN 978-1-5386-3418 -9, DOI: 10.1109/SIELA.2018.8447090.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000454837100038>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85053825957&partnerID=MN8TOARS>

25. **Orozova D.**, E. Sotirova, Modeling of a Learning Management System, IEEE XVIII-th International Symposium on Electrical Apparatus and Technologies SIELA 2014, 2014, Bulgaria, pp. 95-98, ISBN: 978-147995817-7, DOI: 10.1109/SIELA.2014.6871879.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000345744800038>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-84906751001&partnerID=MN8TOARS>

26. **Orozova D.**, M. Georgieva, A model of the process of Big Data with Generalized Net, Electrical Apparatus and Technologies (SIELA), 2016, IEEE 19th International Symposium, Bourgas, 238-241, ISSN: 978-1-4673-9521 -2, ISBN 978-619-160-648-1 , DOI: 10.1109/SIELA.2016.7543030.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000382936800059>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-84985945018&partnerID=MN8TOARS>

27. I. Popchev, **D. Orozova**, S. Stoyanov, IoT and Big Data Analytics in E-Learning, 2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE), IEEE 2019, pp. 1-5, ISBN 978-172816481-6, DOI: 10.1109/BdKCSE48644.2019.9010666.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000563467600027>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85082305981&partnerID=MN8TOARS>

28. **Orozova D.**, I. Popchev, Cyber-Physical-Social Systems for Big Data, IEEE XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, 3-6 June 2020, Bourgas, Bulgaria, pp. 334 – 337, ISBN 978-1-7281-4345-3 , DOI: 10.1109/SIELA49118.2020.9167161.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85091338714&partnerID=MN8TOARS>

29. **Orozova D.**, N. Hristova, Generalized Net Model for dynamic decision making and prognoses, IEEE XX I International Symposium on Electrical Apparatus and Technologies SIELA 20 20 , Bulgaria , 20 20 , pp.330 – 333, ISBN 978-1- 7281-4345-3, DOI: 10.1109/SIELA49118.2020.9167077.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85091339727&partnerID=MN8TOARS>

30. Popchev I., **D. Orozova** , Text Mining in the Domain of Plant Genetic Resources, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems, 2020, pp. 596-600, ISBN 978-172815456-5, DOI: 10.1109/IS48319.2020.9200174.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85092747530&partnerID=MN8TOARS>

31. Hadzhikoleva S., **D. Orozova**, E. Hadzhikolev, N. Andonov, Model of a Centralized System for Quality Assurance in Higher Education, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems 2020, pp. 87-92, ISBN 978-172815456-5, DOI: 10.1109/IS48319.2020.9199951.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85092696472&partnerID=MN8TOARS>

32. Georgieva P., E. Nikolova, **D. Orozova** , Data Cleaning Techniques in Detecting Tendencies in Software Engineering, MIPRO 2020, IEEE 43rd International Convention, 2020, Croatia, Opatija, pp.1272-1277, ISSN: 2623-8764 , DOI : 10.23919/MIPRO48935.2020.9245416.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000790326400188>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85097229443&partnerID=MN8TOARS>

33. **D. Orozova**, S. Hadzhikoleva, E. Hadzhikolev, A Course Gamification Model for the Development of Higher Order Thinking Skills, MIPRO 2021, IEEE 44th International Convention, September 27, 2021 – October 1, 2021, Opatija, Croatia, pp. 1772-1777, ISSN 1847-3946, DOI: 10.23919/MIPRO52101.2021.9597151.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85123052448&partnerID=MN8TOARS>

34. I. Popchev, **D. Orozova**, Data Analytics Approach in Virtual Educational Space, 2021 IEEE Big Data, Knowledge and Control Systems Engineering (BdKCSE), 2021, pp. 1-5, doi: 10.1109/BdKCSE53180.2021.9627247, ISBN 978-166541042-7.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85123994089&partnerID=MN8TOARS>

35. **D. Orozova**, M. Todorova, Applying Reactive Blocks in Courses of Internet of Things, Proceedings of the 10th Annual International Conference of Education, Research and Innovation, Seville, 16-18 November, 2017, pp. 6960-6967, ISSN (print):2340-1095.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000429975307011>

36. **D. Orozova**, M. Todorova, How to Follow Modern Trends in Courses in "Databases" - Introduction of Data Mining Techniques by Example, Proceedings of the 11th Annual International Technology, Education and Development Conference, Valencia, 2017, 8186-8194, ISSN:2340-1079.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000427401303026>

37. Georgieva P., E. Nikolova, **D. Orozova**, Tendencies in Software Engineering Education, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 3601-3609, ISBN: 978-84-09-14755-7.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000530109203108>

38. Georgieva P., E. Nikolova, **D. Orozova**, V. Jecheva, Certification In Teaching Information Technologies For Engineering Graduates, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 4564-4573.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000530109204100>

39. Tuparova D., G. Tuparov, **D. Orozova**, Educational Computer Games and Gamification in the Higher Education – Students' Points of View, 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, Croatia, Opatija, pp.1879-1884 Electronic ISSN: 2623-8764, DOI: 10.23919/MIPRO48935.2020.9245251.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000790326400284>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85097241244&partnerID=MN8TOARS>

40. Magdalena Todorova, **Daniela Orozova**, Generalized Net Model of Sequential Programs, In proc of the 20th International Symposium on Electrical Apparatus and Technologies (SIELA 2018), 3-6 June 2018, Bourgas, Bulgaria, DOI: 10.1109/SIELA.2018.8447068 , ISBN: 978-1-5386-3420-2.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000454837100016>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85053816562&partnerID=MN8TOARS>

41. Zlatev Z., J. Ilieva, D. Orozova, G. Shivacheva, N. Angelova, Design and Research of a Sound-to-RGB smart acoustic device, Multimodal Technologies and Interaction, 2023, 7(8), 79; WoS, Scopus Q2 SJR=0.504 (3a 2022), <https://doi.org/10.3390/mti7080079>

42. **Textbook:** Ivan Popchev, **Daniela Orozova** , Presentation of knowledge in systems with artificial intelligence, Bozich Publishing House, Burgas, 2018, ISBN 978-619-7181-59-3.

BIBLIOGRAPHY

1. Valkanova, V., M. Sandalski, I. Popchev, S. Stoyanov. An educational portal for preparing students for matriculation exams. - Theory-practice interaction: Key problems and solutions. International Scientific Conference, 2011, 287-294. ISBN: 978-954-9370-80-5.
2. NEAA: The National Assessment and Accreditation Agency <https://www.neaa.government.bg/>
3. Stoyanov S., D. Orozova, I. Popchev, E. Doychev Virtual space for continuing education, BSU Scientific conference with international participation "Horizons in the development of human resources and knowledge" volume 2, 2015, pp. 419-425.
4. Todorova M., Approaches, programming environments and languages for checking the correctness of programs and their application in the training of software specialists, Sofia, 2013.
5. 12 Best Web Analytics Tools and Software for Startups <https://squeezegrowth.com/bg/best-web-analytics-tools-for-startups/>
6. Ahlers, R., Garris, R., & Driskell, J. (2014). Games, Motivation, and Learning: A Research and Practice Model. Computer Science Applications. Florida Maxima Corporation.
7. Anderson LW, DR Krathwohl (Eds). A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives: Complete edition, New York, Longman, 2001.
8. Atanassov K. , On Generalized Nets Theory "Prof. M. Drinov" Academic Publishing House, 2007.
9. Atanassov, K., Generalized Nets and Intuitionistic Fuzziness in Data Mining, "Prof. M. Drinov" Publishing House of Bulgarian Academy of Sciences, Sofia, 2020, ISBN 978-619-245-022-9.
10. Barr A., Feigenbaum E., The Handbook of Artificial Intelligence, vol.2, 1984 .
11. Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Retrieved from <http://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf> , last access: 01.02.2023
12. Big Data Definition, Teradata, <https://www.teradata.com/Glossary/What-is-Big-Data>
13. Bloom B., M. Engelhart, E. Furst, et. al., "Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain". New York: David McKay Company, 1956.
14. Bradley, J., Atkins, E. Coupled Cyber-Physical System Modeling and Coregulation of a CubeSat, IEEE Transactions on Robotics, 2015, Vol. 31, No. 2, 443-456.
15. Brookhart S., "How to Assess Higher-Order Thinking Skills in Your Classroom", publ. ASCD, 2010, ISBN: 978-1-4166-1048-9.
16. Buckland, B., F. Schreier & TH Winkler. Democratic Governance Challenges of Cyber Security. DCAF Horizon 2015 Working Paper No. 1, Geneva Center for the Democratic Control of Armed Forces.
17. Egger, N. etc., Themes in Science & Technology Education, 2014, Vol. 7 Issue 1, p.3-17, 15p.
18. Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG). (2015). Brussels, Belgium. ISBN: 978-9-08-168672-3.
19. Bower, J., C. Christensen. Disruptive Technologies: Catching the Wave. Harvard Business Review, January-February 1995, pp. 43-53.
20. Brynjolfsson E., A. McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, WW Norton & Company, 2014.
21. Cao L. Data science: a comprehensive overview. ACM Comput Surv (CSUR). 2017;50 (3) pp. 1-42.
22. Chapman A., "Principles and Methods of Data Cleaning - Primary Species and Species-Occurrence Data.", Report for the Global Biodiversity Information Facility, Copenhagen, 2005.
23. Chatti, M, A, Dychhoff, A, L, Schroeder, U., and Thus, H. A reference model for learning analytics, International journal of Technology Enhanced learning, 2012.
24. Chemchem, A., Drias, H. From data mining to knowledge mining: Application to intelligent agents. International Joint Conference on Artificial Intelligence: Expert Systems with Applications, (2014).
25. Clifton, B., 2008: Advanced Web Metrics with Google Analytics, Wiley, New York, USA.
26. Dietrich D., B. Heller, B. Yang , Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data , Published John Wiley & Sons, Inc., 2015, ISBN:978-1-118-87613-8

27. Ekman P., W.V. Friesen, Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 1971, 17(2), 124–129.
28. Escobar-Jeria, VH, Martín-Bautista, MJ, Sánchez, D., Vila, M., 2007: Web Usage Mining Via Fuzzy Logic Techniques. In: Melin, P., Castillo, O., Aguilar, IJ, Kacprzyk, J., Pedrycz, W. (Eds.), 2007: *Lecture Notes In Artificial Intelligence*, Vol. 4529, Springer, New York, USA, pp. 243-252.
29. Estivill-Castro, V. & Yang, JA (2000). Fast and robust general purpose clustering algorithm. *Pacific Rim International Conference on Artificial Intelligence* (pp. 208-218).
30. Etheridge, D., *Excel Programming: Your Visual Blueprint for Creating Interactive Spreadsheets*, John Wiley & Sons, Inc. 2010.
31. Ethics Guidelines for Trustworthy AI [Online]. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> last access: 03/01/2023
32. Fraley C., Raftery AE (1998). How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. Report No. 329. Department of Statistics University of Washington.
33. Georgieva P., E. Nikolova, D. Orozova, Data Cleaning Techniques in Detecting Tendencies in Software Engineering, MIPRO 2020, 43rd International Convention, 2020, Croatia, Opatija, pp.1272-1277, Electronic ISSN: 2623-8764
34. Georgieva P., E. Nikolova, D. Orozova, Tendencies in Software Engineering Education, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 3601- 3609, ISBN: 978-84-09-14755-7
35. Georgieva P., E. Nikolova, D. Orozova, V. Jecheva, Certification In Teaching Information Technologies For Engineering Graduates, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 4564-4573
36. Gomez-Perez A., M. Fernando-Lopez, O. Corcho. "Ontological Engineering with examples from the areas of Knowledge management," *e-Commerce and the Semantic Web 2nd. Ed.* London, Springer-Verlag, 2004.
37. Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics, *Educational Technology & Society* , 15(3), 42-57.
38. Gruber R., Translation Approach to Portable Ontologies, *Knowledge Acquisition*, 1993 V.5. 199-220.
39. Hadzhikolev, E., S. Hadzhikoleva, K. Yotov, D. Orozova, Models for Multicomponent Fuzzy Evaluation, with a Focus on the Assessment of Higher-Order Thinking Skills, *TEM Journal*, Vol.9, No.4, pp.22 -28, 2020.
40. Hadzhikolev E., S. Hadzhikoleva, D. Orozova, Digital Model of a Document in a University Document Repository, XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018
41. Hadzhikoleva S., D. Orozova, N. Andonov, E. Hadzhikolev, Generalized Net Model of a System for Quality Assurance in Higher Education, 45th International Conference Applications of Mathematics in Engineering and Economics (AMEE'19), 2019 , pp. 040005-1 -040005-8, ISBN: 978-0-7354-1919-3. AIP Conference Proceedings.
42. Hadzhikoleva S., D. Orozova, E. Hadzhikolev, N. Andonov, Conceptual Model of a Centralized System for Quality Assurance in Higher Education, *Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems 2020*, pp. 87-92, ISBN 978-172815456-5, DOI 10.1109/IS48319.2020.9200174
43. Hadzhikolev E., S. Hadzhikoleva, D. Orozova, K. Yotov, A Comprehensive Approach to Assessing Higher and Lower Order Thinking Skills, Uncertainty and imprecision in Decision Making and Decision Support: New Advances, Challenges, and Perspectives, *Lecture Notes in Networks and Systems*, Vol. 338, pp. 164–177, *Lecture Notes in Networks and Systems*, Volume 338, pp. 164–177.
44. Hayashi, C. (1998). What is Data Science? Fundamental Concepts and a Heuristic Example. In: Hayashi, C., Yajima, K., Bock, HH., Ohsumi, N., Tanaka, Y., Baba, Y. (eds) *Data Science, Classification, and Related Methods*. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Tokyo.

45. Han J., Kamber M., Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2001.
46. Hornick M., E. Marcade, S. Venkayala, Java data mining: strategy, standard and practice, A practical Guide for Architecture, Design, and Implementation, 2006 .
47. International Educational Data Mining Society . 2022, <https://educationaldatamining.org> . last access: 03/01/2023
48. Jecheva V., D. Orozova, Ontology-Based Electronic Test Result Evaluation, Advances in Intelligent and Soft Computing, Third International Conference of Software, Services and Semantic Technologies S3T, Springer, 213-214, 2011. ISSN: 1867-5662.
49. Kapp, K. (2012). The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education. Pfeiffer.
50. Kaufmann, M., Meier, A., 2009: An Inductive Fuzzy Classification Approach applied to Individual Marketing, In: Proceedings of the 28th North American Fuzzy Information Processing Society Annual Conference, USA.
51. Kaushik, Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity, Wiley, 2010
52. KNIME – Open for Innovation: <https://www.knime.com/> last access: 03/01/2023.
53. Marr B., “Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance”. John Wiley & Sons Ltd, 2015.
54. Millington, Ian, 2006. Artificial Intelligence for Games. Elsevier Inc. Ch.7.6 Reinforcement Learning, 612-628.
55. Mobasher, B., 2007: Web Usage Mining, In: (Liu 2007, pp. 449-483).
56. Orange system [Online]: [https://orange.biolab.si/training/ introduction-to-data-mining /](https://orange.biolab.si/training/introduction-to-data-mining/), access: 01.08.2023
57. Orozov B., D. Orozova, An AI application in E-Learning, CEUR Workshop Proceedings 3061, pp. 76-80, September 27–28, 2021, Plovdiv, Bulgaria, ISSN 1613-0073.
58. Orozov B., D. Orozova, Rule Based System Against Reinforcement Learning, CompSysTech'21: International Conference on Computer Systems and Technologies '21, June 2021, Pages 67–70, ISBN: 978-145038982-2
59. Orozova D., Appropriate E-Test System Selection Model, Comptes rendus de l'Acad'e'mie bulgare des Sciences, 2019, Vol 72, No 6, pp. 811-820, ISSN 1310-1331.
60. Orozova D., SQL User Queries Execution Model, Advances in Intelligent Systems and Computing book series (AISC, volume 1081), Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives, pp.184-189, 2020, ISSN 2194-5357, DOI: 10.1007/978-3-030-47024-1_20
61. Orozova D., I. Popchev, Cyber-Physical-Social Systems for Big Data, XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, Bulgaria, pp. 334 – 337, ISBN 978-1-7281-4345-3
62. Orozova D., K. Atanassov, Generalized Net Model of Processes Related to Big Data, Comptes rendus de l'Académie bulgare des Sciences, book No 12, vol. 71, 2018, pp. 1679 -1686.
63. Orozova D., K. Atanassov, Model of Big Data Map/Reduce Processing, Comptes rendus de l'Acad'e'mie bulgare des Sciences, 2019, Vol 72, No11, pp.1537-1545.
64. Orozova D., M. Todorova, Ontology concept in courses on students, TEM Journal. Volume 7, Issue 3, Pages 693-697, ISSN 2217-8309, DOI: 10.18421/TEM73-29, 2018. ISSN (print):2217-8309 .
65. Orozova D. , M. Krawczak, Project-based Learning Modeled with a Generalized Net, Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Warsaw, Poland, Vol.12, 2015/2016, pp. 68–79, ISBN 978-83-61551-13-3.
66. Orozova D., M. Todorova, Applying Reactive Blocks in Courses of Internet of Things, Proceedings of the 10th Annual International Conference of Education, Research and Innovation, Seville, 16-18 November, 2017, p.6960-6967, ISSN (print):2340-1095
67. Orozova D. , M. Todorova, How to Follow Modern Trends in Courses in "Databases" - Introduction

- of Data Mining Techniques by Example, Proceedings of the 11th Annual International Technology, Education and Development Conference (INTED), Valencia, 2017, p. 8186-8194, ISSN:2340-1079.
68. Orozova D. , K. Atanassov, M. Todorova, Generalized Net Model of the Process of Personalization and Usage of an e-Learning Environment, Proceedings of the Jangjeon Mathematical Society 19 (2016), No. 4, pp. 615 – 624. ISSN:1598-7264, ISSN (online):2508-7916, ISBN:89-87809-15-3.
69. Orozova, S. Hadzhikoleva, E. Hadzhikolev, A Course Gamification Model for the Development of Higher Order Thinking Skills, MIPRO 2021, 44th International Convention, September 27, 2021 – October 1, 202, Opatija, Croatia, pp. 1772-1777, ISSN 1847-3946.
70. Orozova D., E. Sotirova, Modeling of a Learning Management System, XVIII-th International Symposium on Electrical Apparatus and Technologies, Proceedings of digests, SIELA 2014, 29-31 May 2014, Bulgaria, pp. 95-98, ISBN: 978-147995817-7
71. Orozova D., M. Georgieva, A model of the process of Big Data with Generalized Net, Electrical Apparatus and Technologies (SIELA), 2016 19th International Symposium, Bourgas, 238-241, ISSN: 978-1-4673-9521-2 , ISBN 978-619-160-648-1
72. Orozova D., A. Ivanov, Generalized Net Model of Virtual Collaboration Space, XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018, 3-6 June 2018, Bourgas, Bulgaria, ISBN 978-1-5386-3418-9
73. Orozova D., K. Kaloyanova, M. Todorova, Introducing Information Security Concepts and Standards in Higher Education, TEM Journal. Volume 8, Issue 3, Pages 1017-1024, ISSN 2217-8309, DOI: 10.18421/TEM83-46, August 2019.
74. Orozova D., N. Hristova, Generalized Net Model for dynamic decision making and prognoses, XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, 3-6 June 2020, Bourgas, Bulgaria, pp. 330 – 333, ISBN ISBN 978-1-7281-4345-3.
75. Orozova D. , N. Hristova, An application of analytical data research in e-learning system, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering Volume 1031 (2021) 012058, IOP Publishing , Online ISSN: 1757-899X
76. Petri CA, “ Communication with Automaten”. Bonn: Institut fuer Instrumentelle Mathematik, Schriften des IIM Nr.3, 1962. Also, English translation: “Communication with Automata”, New York: Griffiss Air Force Base. Tech. Rep. RADC-TR-65-377, vol.1, Suppl. 1, 1966.
77. Popchev I., D. Orozova, Towards a multistep method for assessment in e-learning of emerging technologies, Cybernetics and Information Technologies, Volume 20, No 3, Sofia, 2020, pp.116-129, Print ISSN: 1311-9702
78. Popchev I., D. Orozova. Towards Big Data Analytics in the E-learning Space. Cybernetics and Information Technologies, Vol. 19(3), 2019, pp. 16-25. ISSN: 1311-9702.
79. Popchev, D. Orozova, S. Stoyanov, IoT and Big Data Analytics in E-Learning, 2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE), 2019, pp. 1-5, ISBN 978-172816481-6
80. Popchev I., D. Orozova, Text Mining in the Domain of Plant Genetic Resources, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems, 2020, pp. 596-600.
81. Popchev I., D. Orozova, DataScience: Experience and Trends, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering, Volume 1031 (2021) 012057, IOP Publishing, Online ISSN: 1757 -899X, Print ISSN: 1757-8981 doi:10.1088/1757-899X/1031/1/012057.
82. Popchev I. , D. Orozova, Data Analytics Approach in Virtual Educational Space, Big Data, Knowledge and Control Systems Engineering (BdKCSE), 2021, pp. 1-5, ISBN 978-166541042-7.
83. Popchev I., D. Orozova, Algorithms for Machine Learning with Orange System, International Journal of Online and Biomedical Engineering (iJOE), Vol. 19, No. 04, 2023, pp. 109 -123, ISSN: 2626-8493,
84. Popchev, I., I. Radeva, Risk Analysis – an Instrument for Technology Selection. Engineering Sciences (2019) LVI (4) 5-20, ISSN 1312-5702, DOI: 10:7546/ Eng.Sci.LVI.19.04.01.
85. Policy and Investment Recommendations for Trustworthy Artificial Intelligence [Online]

- <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence> last access: 03/06/2023
86. Popper, Karl Raimund (1972). Objective knowledge. Oxford, Clarendon Press
 87. RapidMiner - Depth for Data Scientists: <https://rapidminer.com/>
 88. Reinforcement Learning, Searching for optimal policies: <https://www.cs.upc.edu/~mmartin/Ag4-4x.pdf>
 89. Russell, S and P. Norvig. Artificial Intelligence: A Modern Approach, Prentice Hall, Upper Saddle River, New Jersey, 1995.
 90. Siemens, G, and Long, P, Penetrating the fog-analytics in learning and education. Asynchronous Learning Networks, 2011.
 91. Schaughency E., J. Smith, J. van der Meer, and D. Berg, Classical Test Theory and Higher Education: Five Questions, In: Handbook on Measurement, Assessment, and Evaluation in Higher Education, Edited by Charles Secolsky and D. Brian Denison, Publ. Routledge, 2017, ISBN: 1138892157.
 92. Social Analytics 2022, <https://www.gartner.com/en/information-technology/glossary/social-analytics>
 93. Stoyanov, S., I. Popchev, I., E. Doychev, D. Mitev, V. Valkanov, A. Stoyanova-Doycheva, V. Valkanova, I. Minov. DeLC Educational Portal. – Cybernetics and Information Technologies, Vol. 10, 2010, No. 3, 49-69, ISSN: 1311-9702
 94. Stoyanova-Doycheva A., T. Glushkova, S. Stoyanov, D. Orozova, Lifelong Learning Supported by an Intelligent Tourist Guide, 2018, AIP Conference Proceedings, Volume 2048, pp. 020038; ISSN: 0094243X, ISBN: 978-073541774-8, DOI: 10.1063/1.5082056
 95. Todorov, J., V. Valkanov, S. Stoyanov, B. Daskalov, I. Popchev, D. Orozova. Chapter 6: Personal Assistants in a Virtual Education Space, Sgurev V., Jotsov V., Kacprzyk J. (eds) Practical Issues of Intelligent Innovations. Studies in Systems, Decision and Control, 140, Springer International Publishing AG part of Springer Nature, 2018, ISSN 2198-4182, ISSN 2198-4190 (electronic), DOI:10.1007/978-3-319-78437-3, 22, 131-153
 96. Todorov J., E. Doychev, D. Orozova, A. Stoyanova-Doycheva, IoT Multiagent Assistant in Virtual Educational Space, AIP Conference Proceedings, Vol. 2048, 020031; ISSN: 0094243X,
 97. Todorova M., D. Orozova, Generalized Net Model of Sequential Programs, In proc of the 20th International Symposium on Electrical Apparatus and Technologies (SIELA 2018), 2018, Bulgaria, DOI: 10.1109/SIELA.2018.8447068, ISBN: 978-1-5386-3420-2.
 98. Todorova M., D. Orozova, Training Difficulties in Deductive Methods of Verification and Synthesis of Program, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 9, No. 7, 2018, pp. 18-22, ISSN:2158-107X, doi:10.14569/IJACSA.2018.090703.
 99. Tuparova D., G. Tuparov, D. Orozova, Educational Computer Games and Gamification in the Higher Education – Students' Points of View, 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, Croatia, Opatija, pp. 1879-1884 Electronic ISSN: 2623-8764, DOI: 10.23919/MIPRO48935.2020.9245251.
 100. Unsupervised Learning and Data Clustering: <https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a>, last access: 03/06/2023
 101. Venkatram K., Geetha M. A., Review on Big Data & Analytics – Concepts, Philosophy, Process and Applications, Cybernetics and Information Technologies, Vol.17(2), 2017, 3-27, ISSN:1311-9702;
 102. Wang FY., Driving into the future with ITS, IEEE Intelligent System, Vol.21, No.3, 2006, 94–95.
 103. WEKA Waikato Environment for Knowledge Analysis: <http://www.cs.waikato.ac.nz/ml/weka>
 104. Wordclouds - free online generator: www.wordclouds.com/
 105. Zadeh, L., The concept of a linguistic variable and its application to approximate reasoning, Published 1975, Information Sciences, v. 8, pp. 199–249.
 106. Zlatev Z., J. Ilieva, D. Orozova, G. Shivacheva, N. Angelova, Design and Research of a Sound-to-RGB smart acoustic device, Multimodal Technol. Interact. 2023, 7(8), 79.
 107. Zumstein D., M. Kaufmann, A Fuzzy Web Analytics Model for Web Mining, IADIS European Conference on Data Mining, 2009, pp.59-66, ISBN: 978-972-8924-88-1.



АВТОРЕФЕРАТ НА ДИСЕРТАЦИЯ

за присъждане на научна степен “доктор на науките” в професионално
направление

4.6 “Информатика и компютърни науки”

Приложение на науката за данните във виртуалното образователно пространство

Даниела Ананиева Орозова

Научно жури:

Акад. Иван Попчев

Акад. Красимир Атанасов

Чл.-кор. Любка Дуковска

Проф. Веселин Видев

Проф. Радослав Йошинов

Проф. Мария Нишева-Павлова

Проф. Мария Христова



**Институт по информационни и
комуникационни технологии**

Секция „Интелигентни системи”

Съответствия на използваните термини от английски език

Термин от английски език	Пояснение
Big Data	големи данни
Content Management System (CMS)	система за управление на съдържанието
Cyber-Physical Social System (CPSS)	кибер-физическа социална система
Cyber-Physical Social Educational System (CPSES)	кибер-физическа социална образователна система
Data Analytics	анализи на данни
Data Science	наука за данните
Data Mining	извличане на знания от данните
Data Cleaning	почистване на данните
Data Integration	интеграция на данните
Data Reduction	редуциране на данните
Data Transformation	трансформация на данните
E-Learning	електронно обучение
Educational Data Mining	извличане на образователни данни
Emerging technologies	нововъзникващи технологии
Executable models	изпълними модели
Fuzzy sets and systems	размити множества и системи
Generalized Net (GN)	обобщени мрежи (OM)
Internet of Things (IoT)	интернет на нещата
Knowledge based systems	системи, основани на знания
Knowledge representation	представяне на знания
Learning Analytics	учебни аналитики
Learning Management System (LMS)	система за управление на учебния процес
Learning System	учебна система
Machine Learning	машинно самообучение
Multi-agent systems	мултиагентни системи
Natural language processing	обработка на естествен език
Ontologies engineering	онтологично инженерство
Orange Data Mining system	софтуер за машинно обучение и анализ на данни
Outliers	отличителни стойности
Property-based models	модели, основани на свойства
Reinforcement Learning	подсилващо обучение
Recommendation systems	препоръчващи системи
Sharable Content Object Reference Model (SCORM)	спецификация за стандартизирано електронно учебно съдържание
Semantic Web	семантичен уеб
Supervised and Unsupervised learning	контролирано и неконтролирано обучение

УВОД

1. Науката за данните в обучението

Виртуалните образователни пространства [3] са информационни и социални пространства, които интегрират хетерогенни технологии и различни педагогически подходи. Те са среда за доставка на учебни материали и образователни услуги за различни целеви групи, независимо от времето и пространството. Данните натрупвани от работата на образователното пространство постоянно се увеличават. Много страни и университети по целия свят изграждат инфраструктури за анализиране на тези данни. Siemens & Long [90] определят *Data analytics in Education* като "измерване, събиране, анализ и отчитане на данните за учащите се и техните контексти с цел разбиране и оптимизиране на обучението и околната среда, в която се случва." Chatti [23] определя целите на този анализ като: мониторинг, прогнозиране, индивидуализация, намеса в обучаващия процес, оценка и препоръки на обучаемия.

Термините *Big Data*, *Data Analytics* и *Data Mining* описват както самите данни, така и технологиите за събиране, обработка, управление на данните и методите за анализ [24]. *Data Mining* е процеса на търсене на скрити данни и закономерности, предварително неизвестни, нетривиални и практически полезни, необходими за взимане на решения в различни сфери на човешките дейности. *Big Data Analytics* се явява развитие на концепцията *Data Mining*. Също така е и развитие на решаваните задачи, сфери на приложение, източници на данни, методи и технологии на обработка. През 1998 г. Чикио Хаяши [44] въвежда термина *Data science* като нова, интердисциплинарна концепция, с три аспекта: проектиране на данни, събиране и анализ. *Data science* съчетава множество подходи и техники, свързани с анализ на данни, откриване на знания, машинно обучение, изкуствен интелект, програмиране, комуникация др. *Data science* е „сплав“ от различни дисциплини, технологии и средства за анализ на данни и се явява основно средство за откриване и оползотворяване на потенциала на Big Data [26].

Терминът „*Data analysis*“ се отнася до обработката на данни чрез конвенционални (класически статистически, емпирични или логически) теории, технологии и инструменти за извличане на полезна информация и за практически цели. Терминът „*Data analytics*“, от друга страна, се отнася до теориите, технологиите, инструментите и процесите, които позволяват задълбочено разбиране на изследваните данни [21].

Събирането на данни за анализ на обучението се отнася до целия процес и включва всички данни, получени по време на учебните дейности. Това са много и разнообразни типове данни. Международната организация по стандартизация, IMG Global Learning Consortium (IMG Global) класифицира данните, които могат да бъдат събрани и анализирани в областта на образованието в пет вида: данни за учебното съдържание; данни за учебната дейност; оперативни данни; данни, свързани с кариерното развитие; данни за профила на обучаемия. Това е предпоставка за появата на *Learning Analytics* [23], нов клон на анализ на данни, който включва начертани цели и методи от образователни и психологически изследвания.

През 2009 г. Avinash Kaushik публикува книгата "Уеб анализ 2.0", в която дефинира *Digital Analytics* като „непрекъснато анализиране на количествени и качествени данни от уеб пространството, което да подобри онлайн преживяването на потенциалните клиенти и така да доведе до желаните резултати“ [51]. Въведен е подход на разделяне на данните на различен принцип, което им придава специфично значение. Навлизането и популяризирането на социалните мрежи поставя началото и на съвременния социален анализ *Social Analytics*, определен като "наблюдение, анализ, измерване и тълкуване на цифровите взаимодействия и взаимоотношения на хора, теми, идеи и съдържание" [92]. Този анализ изследва ролята на социалните взаимодействия в процеса на учене и тяхното влияние върху напредъка на обучаемите.

Извличането на образователни данни, наричано още *Educational Data Mining*, е

дефинирано като „дисциплина, занимаваща се с разработването на методи за изследване на уникалните и все по-мощни данни, които идват от образователните среди, и използването на тези методи за по-добро разбиране на обучаемите и на средата, в която те учат“ [47]. Интегрирайки средствата за извличане на знания от данните в обучаващите среди, могат да се търсят тенденции в развитието на процесите на електронното обучение и неговото обслужване. Проследявайки процеса на избор и използване на различни средства могат да се изградят оптимални учебни среди с възможности за персонализирано овладяване на ключови знания, умения и компетенции. Могат да предвидят потенциалната загуба на обучаеми, да се определят кои дейности ще се предпочитат и каква ще е тяхната ефективност. Новите информационни технологии допринасят за изследване, анализиране и развиване на обучаващите среди като наблюдават и измерват различни аспекти на виртуалната и физическата среда, в която се провежда обучението.

2. Цел и задачи на дисертационния труд

Целта на изследването са теоретични обобщения на процесите на наблюдение и анализ на данните от динамичното взаимодействие на обекти в обучаваща среда и създаване на методи и модели за решаване на научни или научноприложни проблеми във виртуалното образователно пространство.

За постигане на целта са поставени следните четири **задачи**:

- Анализ на прилагането на средства за извличане на знания в обучаващите пространства и търсене на решения за персонализиране на електронното и дистанционно обучение.
- Създаване на методи за оценяване и прогнозиране на знанията, уменията и компетентностите на обучаемите във виртуалното образователно пространство.
- Създаване на модели, като резултат от теоретични обобщения на процесите на наблюдение и анализ на дейностите на обучаемите, свързани с Big Data Analytics, Data Mining, Web метрики, обобщени мрежи, машинно обучение и размита логика.
- Въвеждане на базови модули и инструменти на науката за данните за решения на научноприложни проблеми в обучението.

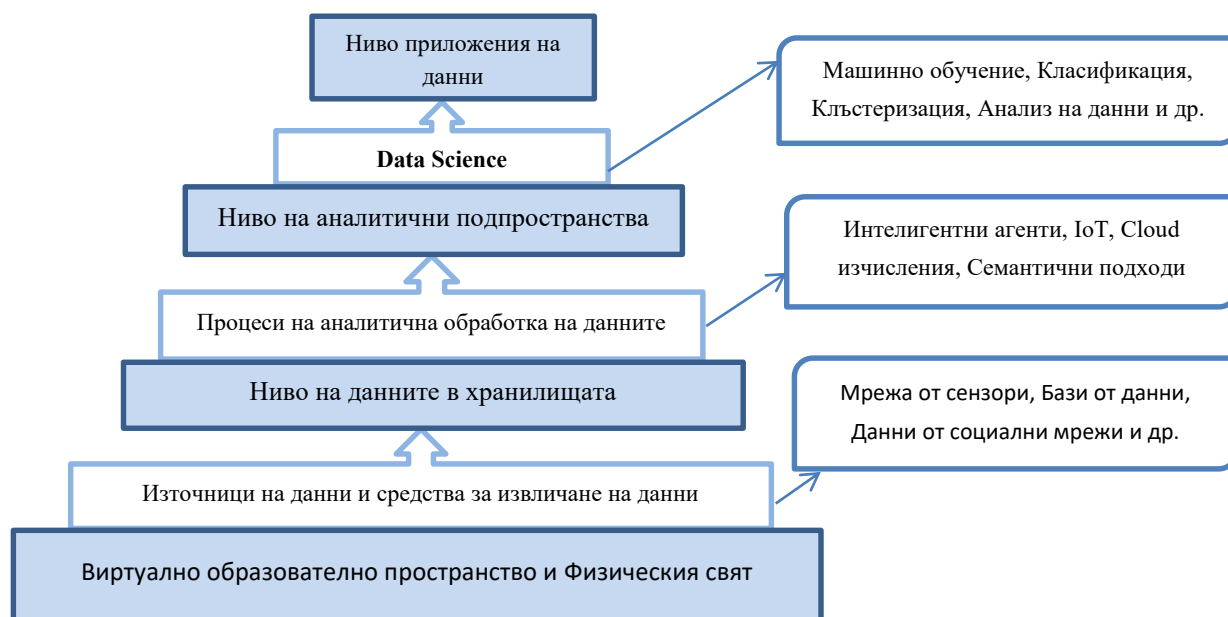
Дисертационният труд на тема „Приложение на науката за данните във виртуалното образователно пространство ” включва 49 фигури, библиографията обхваща 220 източника, от които 197 на английски език.

Текстът е организиран в увод, пет глави и заключение. В **увода** се прави преглед на основните понятия и концепции на изследваната област. Дефинира се целта на дисертационния труд и свързаните с нейното постигане задачи. В **първа глава** се разглежда областта на виртуалното образователно пространство, основните характеристики и възможности на такъв тип инфраструктура и мотивацията за проведената научноизследователска работа. **Втора глава** представя състоянието, предлагани решения, обща характеристика и възможности за интеграция на средства и инструменти към виртуалното образователно пространство. В **трета глава** са предложени методи и техники за моделиране на процесите в образователното пространство, с приложение на инструменти за извличане на данни в образователен контекст (Educational Data Mining). В **четвърта глава** се разглеждат работни процеси във виртуалното образователно пространство и техни формални модели чрез апарата на обобщените мрежи. **Пета глава** представя практики и тенденции при навлизането на науката за данните (DataScience) във висшето образование. **Заключението** обобщава резултатите и дава насоки за продължаване на изследванията по темата на дисертационния труд. Приложени са: Декларация за оригиналност на резултатите, Приноси на дисертационния труд, Списък на публикациите по темата на дисертационния труд.

ГЛАВА 1. ВИРТУАЛНО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО

1.1. Архитектура на данните на виртуално образователно пространство

Разглеждайки архитектурата на виртуалното образователно пространство от гледна точка на данните, акцентът се поставя върху различните нива за представяне на данни и свързаните с тях механизми за събиране и анализ на данни, за извличане на информация и прогнозиране поведението на средата и обучаемите. На фигура 1 е направено обобщение в четири базови нива на представяне на данните, но могат да бъдат разглеждани и други нива в зависимост от прилаганите критерии за съхранение и достъп.



Фигура 1. Функционални нива във виртуалното образователно пространство

Първо ниво: Данни от виртуалното образователно пространство и физическия свят. Източниците обхващат широко поле, обединявайки данни получавани от сензорни датчици, данни от социални мрежи, база от данни за областта на обучението и много други.

Второ ниво: Хранилища на данните. Това ниво предоставя всички данни за състоянието на физическия свят, пренесени във виртуалната среда на пространството. Включва използването на технологии за събиране, съхраняване, обработка на данни и инструменти за анализ.

Трето ниво: Хранилища с аналитични подпространства на данните. Данните във виртуалните образователни пространства могат да бъдат разглеждани от различни гледни точки и подлагани на различни видове анализи.

Четвърто ниво: Приложният слой предоставя на потребителите набор от инструменти и интерфейси за управление, обработка и визуализиране на данни от предходни слоеве.

Новите технологии за работа с големи данни и техническата мощ на съвременните компютри създават възможност за иновативни приложения, базирани на интегриране на разнородни данни.

1.2. Виртуално образователно пространство в концепцията Интернет на нещата

Концепцията за Интернет на нещата (IoT) дефинира устройствата или обектите (нещата), съставляващи екосистема, която обитават и обогатяват чрез споделяне на своите функционалности, съдържание и знания. Електронното обучение може да бъде включено в условията на IoT. Всеки аспект на онлайн образованието е нещо или комбинация от неща в тази екосистема: учене, оценка, интелигентна образователна помощ и др. Навлизането на Интернет

повсеместно в нашия живот и постепенната му трансформация в Интернет на нещата е предпоставка за създаване на “кибер-пространство” [16]. В [14] се дефинира „кибер-физическа система“ (Cyber-Physical-System, CPS) като инженерна система, изградена от изчислителни и физически компоненти. "Физически" са елементите на системата, заемащи физическо пространство, "кибер" са изчислителните и комуникационни елементи на системата. Отчитането на социалната компонента в CPS я трансформира в кибер-физическа социална система (Cyber-Physical Social System, CPSS) [102]. Реализацията в сферата на образованието се явява като CPSES (Cyber-Physical Social Educational System) и представя подмножество на CPSS в домейна „образование и обучение“.

Могат да се посочат успешни резултати за интеграция на виртуалната среда с физическия свят. В лабораторията „Център за електронно обучение DeLC (Distributed eLearning Center)“ на Пловдивския университет се разработва кибер-физическа-социална система, която се изгражда като наследник на системата за електронно обучение DeLC [93]. Тя е наречена VES (Virtual eLearning Space) и се реализира чрез агентно-ориентиран подход. Изгражданото образователното пространство [95] се "населява" от активни автономни компоненти, наречени асистенти. Те са имплементирани като интелигентни софтуерни агенти, които поддържат планирането, организирането и изпълнението на учебния процес [86, 96]. Пространство интегрира виртуалния свят на електронното обучение с интелигентни устройства, сензори за събиране на данни от физическия свят, комуникационни инфраструктури и системна архитектура.

Повечето университети в страната реализират проекти за прилагане на информационни технологии и новаторски подходи, базирани на готови системи за управление на обучението. При тези среди липсва тясна интеграция на виртуалната среда с физическия свят на учебния процес.

В рамките на този дисертационен труд виртуалното образователно пространство се дефинира като среда, която интегрира различни информационни технологии и педагогически подходи за доставка на учебни материали и образователни услуги в абстрактен университет (без да реализира интеграция с физическия свят). Обхваща различни образователни инструменти и ресурси, които се използват в дигитална форма, за да се подобри образователния процес. Усилията са в посока разработване на методология за структуриране и моделиране на процесите на обучение, детайлизиране и формализация на процесите.

1. 3. Виртуалното образователно пространство и големите данни

Виртуалното образователно пространство ежедневно събира огромно количество данни, което изисква нов подход в тяхната обработка. С термина "BigData" се определят усилията в организирането, съхранението, обработката и анализа на големи данни, които са толкова обемни и сложни, че е невъзможна тяхната обработка с традиционните приложения за работа с данни [12]. В [62] се предлага модел, описващ процесите свързани със съхранение, обработка и анализ на големи данни, които изискват нов поглед и съвместно прилагане на редица утвърдени технологии. В [63] се предлага модел на процесите, свързани с разпределени изчисления, базирайки се на Map/Reduce парадигмата, при използване на различни алгоритми и инструменти за реализация.

1. 4. Виртуалното образователно пространство на нововъзникващите технологии и рискове

Водещите технологии се характеризират с радикална новост, бърз растеж и въздействие върху останалите технологии е прието да се наричат нововъзникващи технологии (*emerging technologies*) или подривни технологии (*disruptive technologies*) [19]. В американски източници тези технологии се определят и като *брилянтни технологии* [20], които се развиват във времето и имат потенциал да повишат влиянието си върху процесите на производство на знания. **Изкуственият интелект е „сърцето“** на нововъзникващите технологии, защото свързаните с него научни пробиви формират направления, чието функциониране зависи в най-голяма степен от

представянето на знанията и имитиране на способностите на човешките разсъждения.

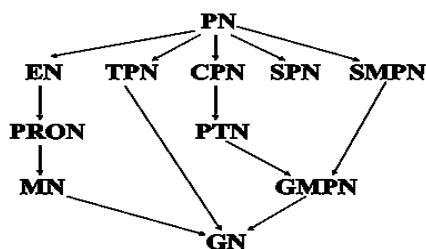
Цифровите технологии стават все по-сложни и интегрирани и като такива те предизвикват значителни трансформации в обществото и икономиката. Тази промяна има **положителни въздействия**, но трябва да отчитаме и **отрицателните въздействия** като например: загуба на работни места в традиционни професии; киберпрестъпност и хакерски атаки; грешки и неразбираемост; необходимост от нови концепции за отговорност, отчетност и управление; повишаване на неравенството между хората. Изложените на най-голям брой рискове са технологиите на „Изкуствен интелект и вземане на решения“, „Големи данни и решения“, „Икономика на споделянето“, „3D печат и производство, потребителски продукти и здравеопазване“, „Нашето цифрово присъствие“ и „Автономни автомобили“ [84]. За да продължи да се развива концепцията на виртуалното образователно пространство в тази динамична и хетерогенна среда, тя трябва да се адаптира към характеристиките и изискванията, които средата налага. Но всички промени трябва да са подчинени на политика и инвестиции за надежден изкуствен интелект [85] и да са основани на етичен и ориентиран към човека подход [31].

1. 5. Модели за анализ на предметната област

В литературата [4] формалните модели за анализ на програмно осигуряване се разделят на: модели, основани на свойства, изпълними модели и модели интегриращи двата подхода.

Към изпълнимите модели се разглеждат мрежите на Петри като просто и ясно средство за моделиране [76]. Те имат разширения, добавящи нови свойства и възможности за моделиране. Независимо от техните спецификации, всички тези инструменти имат позиции и преходи (дъги) и ядра, които се движат в мрежата. Основни разширения на мрежите на Петри [8, 9]:

- o EN – *Evaluation nets* добавят продължителност на движението на ядрото;
- o TPN – *Temporal PN* добавят момент за активиране на прехода;
- o CPN – *Color PN*. Всяко ядро има цвят и може да се движи само по дъги от същия цвят;
- o SPN – *Stochastic PN*. Изборът на дъга се основава на произволно генерирано число;
- o SMPN – *Self Modifying PN*. При самомодифициращите се мрежи на Петри, изборът на дъга се основава на генерирано двоично число 0 или 1, (като при 0 преходът не се извършва);
- o PRON – *Pro-Net* въвежда тип на прехода;
- o PTN – *Predicate/Transition Net*. Тези мрежи дефинират условие на прехода;
- o MN – *M Net*, при които ядрата изпълняват и допълнителни процедури;
- o GMPN – *Generalized Modifying PN* – ядрото се поглъща, когато предикатът генерира 0, в противен случай при генериране на 1 ядрото извършва прехода.
- o GN *Generalized Net* – обединяват различните разширения в единен формализъм за описание на паралелно протичащи процеси. Различните разширения на мрежите на Петри и връзката между тях се представя на фигура 2.



Фигура 2. Разширения на мрежите на Петри

В четвърта глава са представени формални модели на процеси и компоненти на виртуалното образователно пространство чрез апарата на обобщените мрежи. Те са обобщение на създадена серия от модели, свързани с наблюдение и анализ на дейностите на обучаемите в образователна среда, представени в публикациите [59, 62, 63, 68, 40, 41, 70, 71, 72, 74, 69, 97, 99].

ГЛАВА 2. СРЕДСТВА НА ВИРТУАЛНОТО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО

Виртуалното образователно пространство се разглежда като среда, която интегрира различни информационни технологии, педагогически подходи, образователни инструменти и ресурси, които се използват в дигитална форма, за доставка на учебни материали и образователни услуги в абстрактен университет.

Науката за данните може да даде множество ползи в управлението на университет и да помогне за подобряване на ефективността и вземането на решения. Като се имат предвид конкретните нужди и цели на образователната институция могат да се изберат подходящи инструменти и методики за анализ на данните. Пространството се разглежда като активна система, отворена за компоненти с общ подход за достъп и използване на информацията [1, 55]. Такава система осигурява базови компоненти като:

- Средства за събиране, пречистване и ефективно съхранение на данните. Съхраняваната информация да бъде достъпна на различни нива на достъп и от различни приложения за извличане, показване и вземане на решения, както и контрол на достъпа до данните. Възможности за свързване на информацията, получена от различни източници.
- Средства за управление на знания с възможности за интелигентно търсене и четене, контрол за цялостност и коректност на данните, автоматизирано извличане на информация и др. Тук се включат инструменти за управление на качеството на данните, автоматизирано откриване на модели в данните и възможности за самообновяване на данните.
- Средства за оптимизиране на процеса на изпълнение на заявки за търсене и достъп до информацията. Средства за преобразуване и визуализиране на информацията, за по-лесно възприемане и използване от потребителя.
- Възможности за интегриране на компоненти с разнородни технологии и педагогически методи, посредством възприемане на общи архитектурни подходи за системата.

Освен базовите средства, начин за ангажиране на вниманието на обучаемите е да се интегрират в учебния процес онези средства, които те използват ежедневно в личното си виртуално пространство.

Интегрирането на средства за извличане на знания в средите за обучение може значително да подобри образователния процес като предостави ценна информация на обучаемите, преподавателите и администраторите, но от своя страна изисква ефективно управление на образователните данни. Това включва съхранение, обработка и защита на данните, като се осигурява тяхната сигурност и поверителност. Те трябва да бъдат използвани съобразно етични принципи и да зачитат правата на личността и поверителността.

В тази глава от дисертационния труд са обобщени резултати от проведени изследвания и анализи на основни характеристики на виртуалното образователно пространство на Университет, представени в публикациите, свързани с темата на дисертационния труд [61, 79, 82, 94, 96, 97]. Обосновани са предимства на стандартизираното електронно учебно съдържание при доставка на образователни услуги. Мотивирана е необходимостта от прилагане на средства за извличане на знания от данните в обучаващите пространства и търсене на решения за персонализиране на електронното и дистанционно обучение.

ГЛАВА 3. МОДЕЛИ ЗА АНАЛИЗ НА ДАННИ ВЪВ ВИРТУАЛНОТО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО

В тази глава от дисертационния труд са обобщени резултатите от проведените експерименти и анализи на събраните данни за обучаемите в различни форми на електронно обучение, подробно представени в авторските публикации [75, 79, 80, 82]. В проведените изследвания се търси връзка между статистически методи, машинно обучение, откриване на поведенчески модели и анализ на данни. Разработените алгоритми и софтуер са представени в [57, 58, 83]. Предложени са модели, като резултат от теоретични обобщения на процесите на наблюдение и анализ на дейностите на обучаемите, свързани с Big Data Analytics, Data Mining, уеб метрики, машинно обучение и размита логика.

Анализи на данни във виртуално образователно пространство

За оценяване на знанията на обучаемите се използват различни форми и методи на изпитване като: online или offline тестове, отговори на отворени въпроси, тестване и разчитане на код, решаване на задачи и др. Крайната оценка по дисциплината често е комплексна и включва няколко компонента, участващи във формирането ѝ с различни тегла. Тя трябва да отразява различните страни от подготовката на обучавания (теоретични знания и практически умения), да е съобразена със спецификата на оценяването, с възрастови и индивидуални особености и др.

Експеримент за оценяване и аналитични изследвания на данните за обучаемите, е представен в [75]. За целите на анализа разглеждаме конкретен подход за оценяване на обучаемите, при който оценката в края на семестъра се определя като претеглена средна стойност (weighted mean \bar{x}) от всички оценки през семестъра. Всяка оценка има важност или тегло (weight). Използваната в конкретния случай функция има вида:

$$(1) \quad \bar{x} = \frac{\sum xw}{\sum w}$$

където x е оценката, w е теглото на всяка оценка.

Оценката на студентите по предмета Математика в края на семестъра се определя въз основа на следните изпитвания: Тест 1 Множества (*Test 1 Sets*), Тест 2 Логика (*Test 2 Logic*), Тест 3 Геометрия (*Test 3 Geometry*), Тест 4 Статистика (*Test 4 Statistics*), Тест 5 Вероятности (*Test 5 Probability*), изпит в края на семестъра (*Final Exam*), контролни (*Quizzes*), домашна работа (*Homework*), проекти (*Projects*) и работа в часа (*Class Activities*). Всеки критерий има степен на важност и функцията има вида:

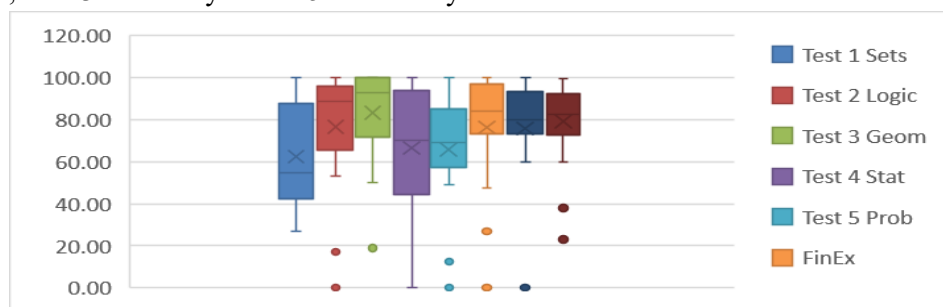
$$(2) \quad \bar{x} = \frac{\sum T1*0.09+T2*0.09+T3*0.09+T4*0.09+T5*0.09+FinEx*0.15+Act*0.40}{\sum w}$$

където \bar{x} е оценката в края на семестъра, w е теглото на всяка оценка. Разглеждаме как получената оценка по отделните тестове влияе на оценяването и как дейностите в часа и извън часа се отразяват на оценката в края на семестъра. Освен точките, натрупвани по отделните компоненти, съхраняваме и данни за всеки студент: Age (години е студент), Gender (пол), New Student Experience (NSE), FullTime-PartTime (кредитите за предмета са 12 или повече), Student Program (тип на обучението), Times Taken Course (за кой път студентът изучава предмета).

3.1. Използване на електронни таблици за аналитични изследвания на данните

Електронните таблици (в частност Excel) предлагат ефективни функции за обобщаване на данните. Филтрите (*Filters*) дават възможности за подбор на записи. Условното форматиране (*Conditional formatting*) оцветява данните, които отговарят на определени критерии и ни помага да откриваме отклонения и тенденции в данните. Графиките (*Charts*) визуално показват отклонения и тенденции [30]. Използвайки инструментите на Excel се извърши първоначален анализ на резултатите от представянето на студентите по математика в продължение на пет семестъра. Графиката от фиг. 3 показва, добро усвояване на учебния материал. Най-ниски са оценките по Test

4 Statistics, където 75% от обучаемите имат оценки по-ниски от 50% от обучаемите по Test 1 Sets, Test 2 Logic, Test 3 Geometry и Test 5 Probability.



Фигура 3. Графика, относно разпределението на оценките по модули

В резултат на изследването бе направена препоръка за промяна на последователността в изучаването на две дисциплини. Промяната на стратегиите за преподаване оказва влияние върху начина, по който обучаемите усвояват материала.

3.2. Аналитични изследвания на данните в обучаващото пространство със системата „Orange“

Достъпни са различни инструменти за извличане на данни. Сред най-популярните са: RapidMiner, RapidAnalytics, WEKA, PSPP, KNIME, Orange, Apache Mahout, jHerWork, Rattle, GhostMiner, XENO, SAS Miner, Polyanalyst, IBM SPSS modeler. За целите на това изследване е разработен инструмент за идентифициране и прогнозиране на причините за изоставане или отпадане на обучаеми, представен в [78]. Използвана е системата „Orange Data Mining system“ [56], софтуер с отворен код за машинно обучение и извличане на данни, написан на Python.

Контролирано обучение

Основната задача при контролираното обучение (*Supervised learning*) е да се създаде модел от обозначени (етикирани) данни, което позволява да се правят прогнози за бъдещи данни. Основни техники са: класификация, когато етикетите на класовете са дискретни и регресия, когато резултатът е непрекъснатата стойност. В системата *Orange Data Mining System* са създадени редица инструменти за изграждане на модели за класификация и регресия. Следва представяне на експерименти за оценяване и прогнозиране на знанията на обучаемите в образователно пространство с прилагане на инструментите: Logistic Regression, Naïve Bayes Classifier, Support Vector Machines (SVM), Decision Trees and Artificial Neural Networks и др.

3.2.1. Проблеми при подготовка на данните за анализ и подходи за решаване

За анализ на данните в образователно пространство, основни входни данни са точките получени от обучаемите на различни оценъчни компоненти, а изходни са съответните крайни оценки, поставени от преподавател. Основни задачи за подготовка на данните са [22]: почистване на данните (Data Cleaning), интеграция на данните (Data Integration), трансформация (Data Transformation), редуциране на данните (Data Reduction). Почистването на данните е процесът за гарантиране, че данните са правилни, последователни и използваеми. *Дублиращи* се данни могат да възникнат при комбиниране на набори от данни от множество източници. *Неподходящи* данни са такива, които не отговарят на решавания проблем. *Структурни грешки* в данните могат да възникнат по време на измерване или прехвърляне на данни. *Outliers* (отличителни стойности) могат да причинят проблеми с някои видове модели. Когато не разполагаме с пълното множество от данни за дадена характеристика имаме *липсващи данни* (Missing Data). Процесът на подготовка на данни, включва идентифициране на грешки, коригиране, изтриване, подредба или друга обработка, като се запазят само потенциално полезните данни.

3.2.2 Оценка и прогнозиране на знанията на обучаемите във виртуалното образователно пространство

Тук се предлага като **многостъпков процес**, метод за оценяване и прогнозиране на знанията, уменията и компетентностите на обучаемите във виртуалното образователно пространство [66]. **Методът включва пет стъпки** и може да бъде адаптиран към различни курсове на *Emerging technologies* във виртуалното образователно пространство.



Фигура 4. Многостъпков процес на анализ и прогнозиране на знанията на обучаемите

Стъпка 1. Избор на начин за оценяване. Определяне на ключови знания и компетентности за изучаваната технология. Определяне на степен на тежест (важност) на всяка анализирана компетентност.

- Основните **теоретични знания** се оценяват с помощта на компоненти като: *междинни тестове, решаване на задачи и казуси, изпити, обобщена дискусия* и др. Тези компоненти оценяват придобитите знания и способностите за разбиране на изучавания теоретичен материал. Използват се тестови въпроси, изброяване и сравняване на обекти, примери за понятия и използване на алгоритми, въпроси, свързани с обясняване и визуализиране на решения.

- Основните **компетентности** се оценяват от възможностите на студентите да прилагат придобитите знания за вземане на нестандартни решения в: *контролни и домашни работи, курсови задачи и проекти*, в които се решават нови задачи, извършва се критичен анализ на решения, определят се потенциални рискове, правят се самостоятелни изводи и заключения.

Стъпка 2. Процес на оценяване. По време на цялостния процес на обучение се провеждат и процедурите за оценяване. Резултатите се натрупват и съхраняват с цел профилиране на студентите, последващи анализи и прогнозиране в оценяването на нови студенти. В получените резултати се търсят зависимости в отделните оценъчни компоненти между теоретичните и практическите знания, умения и компетентности. В някои случаи студент получава висока оценка на тестовете и слаба оценка на контролната работа или обратно. Създаването на автоматизиран алгоритъм за оценяване при налични специфични стойности на оценъчните компоненти изисква нестандартно решение.

Стъпка 3. Анализ на натрупаните данни от проведените обучения в реална среда. Налични са различни алгоритми за машинно обучение. Основни входни данни са точките, получени от оценъчните компоненти, а изходни са съответните крайни оценки на студентите по дисциплината. При методите за обучение, част от примерните данни се ползват за обучение, а друга – за тестване. При недобри резултати от тестването, процесът на обучение може да се повтори или да се прецени, че избраният подход не е добър за решение на проблема.

Стъпка 4. Създаване на модел за предвиждане. На базата на натрупаните данни се създава класификатор, който извършва прогнози за текущи оценки по дисциплината. Системите за анализ на данни като: SPSS, Orange, Weka и други предлагат средства, на базата на алгоритми за: дърво на решенията, логистична регресия, теорема на Бейс, невронни мрежи и др. След приключване на обучението, моделите могат да се прилагат върху нови данни.

Стъпка 5. Оценка на точността на предвиждане на алгоритмите. След обучението се извършва тестване за точност и прецизност на работата на модела.

Доколко е добър един класификатор се определя от стойностите на съответни метрики за оценка на качеството. Не е достатъчно да се гледа само една от тях. Трябва да се вземат под внимание няколко метрики, като избора на това кои са по-важни е в зависимост от задачите и съответните цели.

Confusion Matrix е един от най-популярните начини за оценка на качеството на класификацията [89]. Представява матрица с размер $N \times N$, където N е броят класове на целевата променлива. Чрез прилагане на този инструмент могат да се съпоставят определените от класификатора и действителните стойности по 4 метрики: True Positive (*TP*), False Positive (*FP*), False Negative (*FN*) и True Negative (*TN*).

- *Общата точност* е показател, който дава информация за това каква част от всички случаи са правилно класифицирани:

$$(3) \quad Accuracy (ACC) = (TP+TN) / (TP+FP+FN+TN)$$

Общият брой правилно класифицирани обекти се разделя на всички случаи.

- *Общата грешка* показва каква част от всички обекти класификаторът е разпределил в грешните класове:

$$(4) \quad ERR = (FP+FN) / (TP+FP+FN+TN) = 1 - ACC$$

Може да се изчисли като общият брой грешно класифицирани обекти се раздели на всички случаи или ако от 1 се извади оценката *Accuracy*.

Нужно е да се вземат под внимание и метриците: *Precision*, *Recall*, *Specificity* и *F1-score*. Тези метрики са най-популярни и често използвани при оценка на класификатор:

- *Precision* - метриката показва каква част от обектите, класифицирани като положителни са в действителност положителни:

$$(5) \quad Precision = TP/(TP+FP)$$

- *Recall* или True Positive Rate (TPR) - дава информация за това каква част от положителния клас е била открита от класификатора:

$$(6) \quad Recall = TP/(TP+FN)$$

- *Specificity* или True Negative Rate (TNR) - показва каква част от отрицателния клас е била намерена от класификатора:

$$(7) \quad Specificity = TN/(TN+FP)$$

- *F1-score* - обобщава *Precision* и *Recall* в единствена стойност. Стойността на този показател е максимална, когато *Precision* и *Recall* са равни:

$$(8) \quad F1-score = 2*(Recall*Precision)/(Recall+ Precision)$$

Анализът на резултатите е свързан с решения от страна на ръководителите на обучаващия процес, как да се стимулират обучаемите и как да се помага на обучаемите в **риск**: допълнителни упражнения и задачи, допълнителна индивидуална и екипна работа, обяснения стъпка по стъпка, работа със софтуер като инструмент за моделиране и т.н.

Приложение на метода за оценяване. Изследвания и анализи на метода за оценяване и прогнозиране са извършени по дисциплината „**Изкуствен интелект**“ с използване на инструментите на *Orange system* за експериментирание и изводи [66]. Представяната дисциплина (с преподаватели акад. Иван Попчев и проф. д-р Даниела Орозова) е част от виртуалното образователно пространство. Тя е структурирана в четири модула: Изкуствен интелект – характеристики и проблеми; Търсене на решение в пространството на състоянията; Представяне на знания; Интелигентно вземане на решения. В средата за електронно обучение по дисциплината има предоставени материали с обем 587 МВ текстово описание, примери и 14 линка с полезни връзки. В преподаването се използват средите Moodle и Microsoft Teams. Тази дисциплина е

основа на курсовете: „Анализ и проектиране на бази от данни и знания“, „Управление на знания в компютърни системи“ и др.

Съпка 1. За оценяване на студентите по дисциплината “Изкуствен интелект” са определени оценъчни компоненти, всеки, от които инспектира теоретични знания, практически умения и компетентности с различно познавателно ниво. Дефинирани са 3 теста (3 x 5 = 15 точки), Контролна работа quiz (15 точки), Проект - 45 точки и Обобщена дискусия - 25 точки.

В **проекта**, всеки студент сам избира тема от основните направления на изкуствения интелект, като например: Ontologies engineering, Semantic Web, Knowledge representation, Computational intelligence, Robotics, Natural language processing, Machine Learning, Deep learning, Soft computing, Pattern recognition, Multi-agent systems, Artificial neural networks, Genetic algorithms, Knowledge based systems, Decision support systems, Business intelligence, Data Science, Fuzzy sets and systems, E-learning и т.н. Проектът задължително включва: състояние, тенденции на развитие, идентификация, анализ и оценка на избрания инструментариум за въздействие върху потенциалните рискове, нерешени проблеми, изводи, заключение и библиография.

Обобщената дискусия със студента е върху темата на проекта, риск мениджмънт, мониторинг, контрол и оценка на риск мениджмънта на потенциалните рискове и възможностите за решаване на нови задачи с нестандартни решения.

Използваният примерен модел за оценяване по дисциплината е представен на фиг.5. Този модел на оценяване може да бъде динамично променян и адаптиран към конкретната дисциплина. Например в дисциплината „**Управление на знания в компютърни системи**“ се прилага алтернативен модел.

Оценъчен компонент	Означение	Оценъчна скала
Теоретични знания модул 2	Тест_1	до 5 точки
Теоретични знания модул 3	Тест_2	до 5 точки
Теоретични знания модул 4	Тест_3	до 5 точки
Практически компетентности	Контролна работа	до 15 точки
Практически знания и умения	Проект	до 45 точки
Теоретични знания и умения	Обобщена дискусия	до 25 точки
Крайна оценка	Окончателна оценка	Точките в оценка [2, 6]

Фигура 5. Примерен модел за оценяване по дисциплината

Стъпка 2. *Провеждане на обучение и процес на оценяване.* Задачата е свързана с намиране на общ подход за автоматизирано оценяване и прогнозиране на оценките на обучаемите. За намаляване субективността при оценяването на практическите умения се допуска оценяването да се извършва от външен оценител от фирми в областта на информационните технологии като: Технологика, Скейл Фокус и др. По всяка дисциплина студентите могат да получат до 100 точки и окончателната оценка се формира по следната скала: от 54 до 60 точки - *Среден* (3); от 61 до 70 точки - *Добър* (4); от 71 до 80 точки - *Много добър* (5); от 81 до 100 точки - *Отличен* (6).

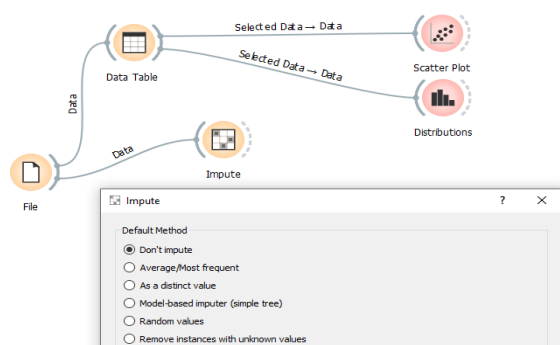
Стъпка 3. Анализ на натрупаните данни от проведените обучения в реална среда.

В процеса на работа са проведени множество експерименти. Основната цел е да се решава класификационен проблем като се установи дали е възможно да се предвиди оценката (изходна променлива) с помощта на входните променливи (точките от отделните оценъчни компоненти), които се запазват в модела. За решаване на класификационния проблем се прилагат различни техники, чрез инструментите на *Orange Data Mining System*.

В началото се създава работен процес и чрез инструмента „File“ **зареждаме данните** за оценките на студентите по различните оценъчни компоненти в брой точки. Те могат да бъдат въведени от Excel (.xlsx), от текстов файл с раздели (.txt), файл с данни разделени със запетая (.csv)

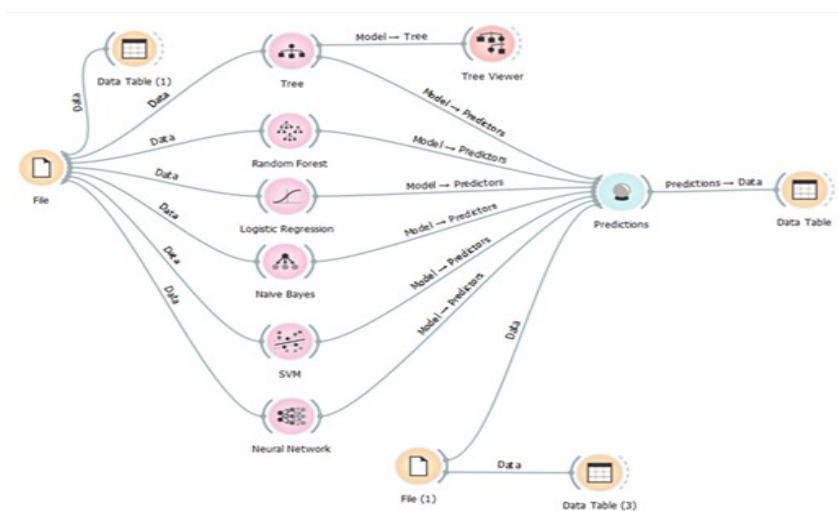
или URL адрес. За по-доброто разбиране, данните могат да се визуализират по някои колони или извадки от тях. Например, може да се свърже файла с данни с инструмента „Scatter Plot“ и се избераат колоните, чиито стойности ще се изобразяват по осите X и Y, използвани цветове, форми, размери и други параметри. Друг популярен инструмент за визуализиране на данните е „Distribution“, чрез който може да се покаже разпределение в набора от данни по даден атрибут.

В зависимост от целта и вида на данните избираме конкретен инструмент за регресия или класификация и **задаваме целевата променлива**. Следва **почистване на данните**. Могат да се приложат различни подходи: изтриване на липсващите стойности или да се заместят по подходящ начин. Чрез инструмент „Impute“ (фиг. 6), се избира измежду различни методи за заместване. По подразбиране е опцията „Remove the rows with missing values“. Други възможни опции са: *Distinct Value, Random Values, Model-Based*.



Фигура 6. Прилагане на инструменти за зареждане и почистване на данните

Стъпка 4. Създаване и трениране на модели за предвиждане. В експеримента последователно се прилагат инструментите на системата: *Tree, Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machines (SVM) и Neural Network*. Работния поток за създаване и трениране на моделите е показан на фиг. 7.



Фигура 7. Работен поток за трениране на моделите и предвиждане

Предвиждане на нови данни. Чрез обучените алгоритми се очаква при задаване на нова комбинация от стойности за избраните компоненти да се определя изходната променлива. На тази стъпка считаме, че моделът е готов за практическо приложение. Моделът придобива независимост и прави свои заключения въз основа на набори от данни и обучение. На фигура 7 е представен работния поток за предвиждане чрез инструмента *Prediction* на системата Orange. Новите данни се подават чрез файл *Test.xlsx*, който има същата структура като началната таблица с данни, но колоната с изходната променлива не е зададена. Общ вид на получавания резултат от

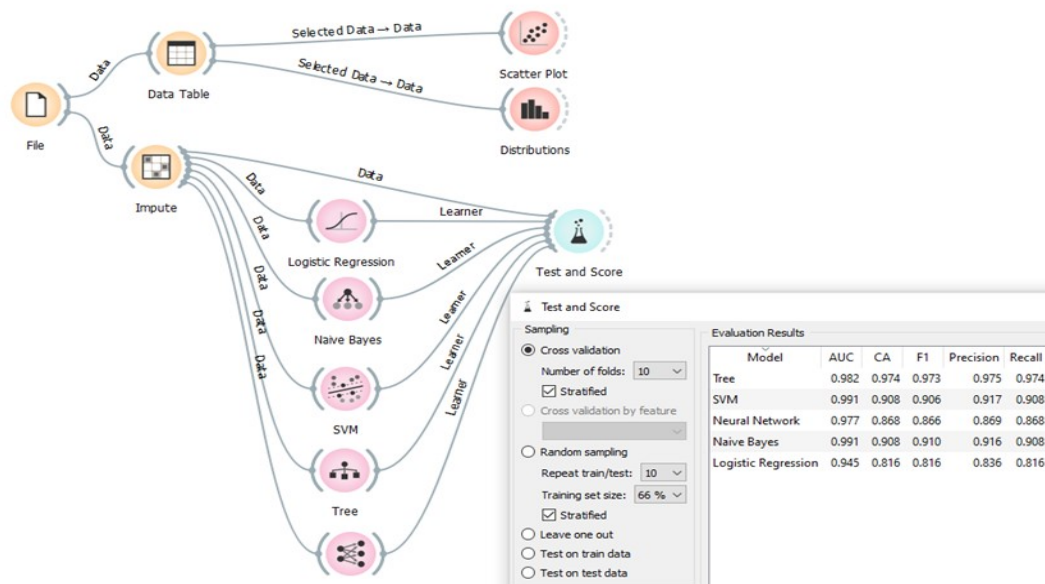
предвиждането е показано на фигура 8. Това са резултатите от предвижданията за крайната оценка на обучаемите, получени от различните модели.

Data & Predictions										
Tree	Logistic Regression			Random Forest			SVM		Naive Bayes	
1 0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Excellent	0.00 : 0.49 : 0.02 : 0.06 : 0.43 → Excellent	0.00 : 0.57 : 0.11 : 0.00 : 0.32 → Excellent	0.01 : 0.18 : 0.03 : 0.01 : 0.77 → Very Good	0.00 : 0.85 : 0.01 :						
2 0.25 : 0.00 : 0.50 : 0.25 : 0.00 → Good	0.01 : 0.00 : 0.45 : 0.22 : 0.32 → Good	0.20 : 0.10 : 0.20 : 0.50 : 0.00 → Middle	0.22 : 0.02 : 0.23 : 0.51 : 0.03 → Middle	0.39 : 0.00 : 0.04 :						
3 0.25 : 0.00 : 0.50 : 0.25 : 0.00 → Good	0.00 : 0.00 : 0.20 : 0.30 : 0.50 → Very Good	0.17 : 0.00 : 0.76 : 0.05 : 0.02 → Good	0.03 : 0.05 : 0.42 : 0.08 : 0.42 → Very Good	0.57 : 0.00 : 0.22 :						
4 0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Excellent	0.00 : 0.72 : 0.03 : 0.04 : 0.22 → Excellent	0.00 : 0.60 : 0.20 : 0.00 : 0.19 → Excellent	0.01 : 0.43 : 0.03 : 0.01 : 0.52 → Very Good	0.00 : 0.25 : 0.04 :						
5 0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Excellent	0.00 : 0.03 : 0.31 : 0.37 : 0.29 → Middle	0.00 : 0.20 : 0.44 : 0.02 : 0.33 → Good	0.03 : 0.06 : 0.58 : 0.08 : 0.24 → Good	0.01 : 0.04 : 0.86 :						
6 1.00 : 0.00 : 0.00 : 0.00 : 0.00 → Bad	0.64 : 0.00 : 0.07 : 0.28 : 0.01 → Bad	0.92 : 0.00 : 0.00 : 0.08 : 0.00 → Bad	0.87 : 0.03 : 0.02 : 0.06 : 0.02 → Bad	0.86 : 0.00 : 0.00 :						
7 0.00 : 0.00 : 0.00 : 0.00 : 1.00 → Very Good	0.00 : 0.00 : 0.00 : 0.20 : 0.79 → Very Good	0.10 : 0.15 : 0.00 : 0.00 : 0.75 → Very Good	0.02 : 0.20 : 0.04 : 0.03 : 0.71 → Very Good	0.01 : 0.60 : 0.00 :						
8 0.00 : 1.00 : 0.00 : 0.00 : 0.00 → Excellent	0.00 : 0.80 : 0.01 : 0.01 : 0.18 → Excellent	0.00 : 1.00 : 0.00 : 0.00 : 0.00 → Excellent	0.02 : 0.87 : 0.03 : 0.03 : 0.06 → Excellent	0.00 : 1.00 : 0.00 :						
9 1.00 : 0.00 : 0.00 : 0.00 : 0.00 → Bad	0.11 : 0.00 : 0.01 : 0.08 : 0.79 → Very Good	0.47 : 0.25 : 0.14 : 0.13 : 0.00 → Bad	0.29 : 0.09 : 0.27 : 0.25 : 0.10 → Good	0.46 : 0.16 : 0.33 :						

Фигура 8. Резултати от предвиждането чрез различни модели

Стъпка 5. Оценяване на работата на моделите, върху данните.

В работния поток се свързва всеки от създадените модели с “Test and Score” инструмент. След като се оценят моделите, трябва да се види дали тяхната точност може да бъде подобрена, чрез настройка на параметрите, присъстващи във модела.



Фигура 9: Работен поток за оценяване с инструмент “Test and Score”

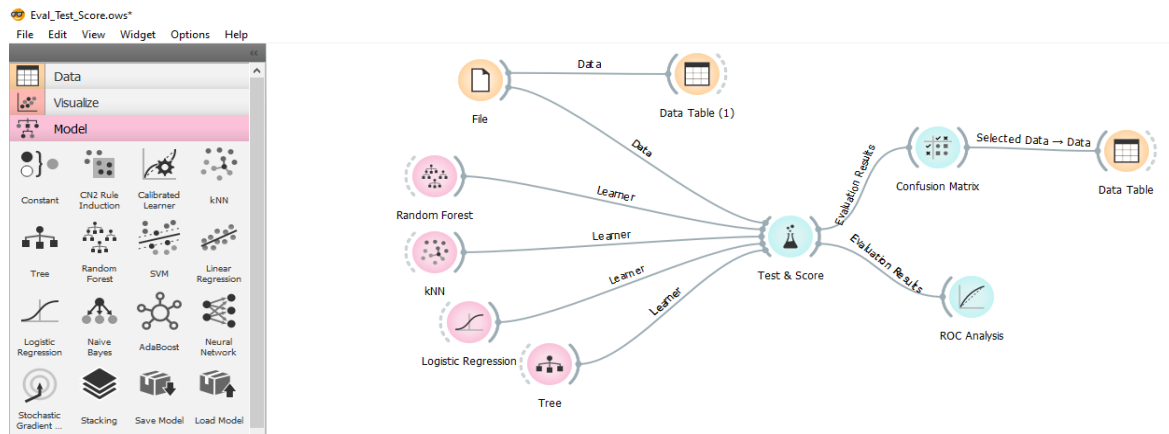
Резултат от работата на инструмент “Test and Score” е таблица с оценки за: Accuracy, Precision, Recall и F1 Score за създадените модели. Конкретни оценки за метриците за оценка на качеството за създадените модели в експеримента са дадени на фиг. 10.

Scores						
Method	AUC	CA	F1	Precision	Recall	
Tree	0.965	0.855	0.884	0.864	0.905	
Logistic Regression	1.000	0.776	0.976	1.000	0.952	
Random Forest	1.000	0.895	0.930	0.909	0.952	
SVM	0.975	0.842	0.900	0.947	0.857	
Naive Bayes	0.994	0.816	0.952	0.952	0.952	
Neural Network	0.941	0.803	0.864	0.826	0.905	

Фигура 10. Резултат от работата на инструмента Test&Score върху създадените модели

Разглеждайки прогнозната точност за всеки един от класовете оценки, може да се обобщи, че тя е най-лоша за класа оценки Middle (оценка Среден). Най-висока точност се постига за класовете Bad (Слаб) и Excellent (Отличен). Прогнозите за оценка Good (Добър) и Very Good (Много добър) при всички разгледани модели се представят с точност около 60-75%. Моделът на Random Forest е най-надежден, защото се представя с най-висока точност за всички класове оценки. Моделът Naive Bayes е с най-ниска оценка за точността спрямо другите модели при разглежданите данни.

Тестване на различните модели и визуализация на точността на работата на моделите може да се извърши чрез „Confusion Matrix”, ROC Analysis или друг инструмент на системата. Работният поток е показан на фиг.11.



Фигура 11. Работен поток за създаване на модели и оценка с чрез Confusion Matrix и ROC Analysis

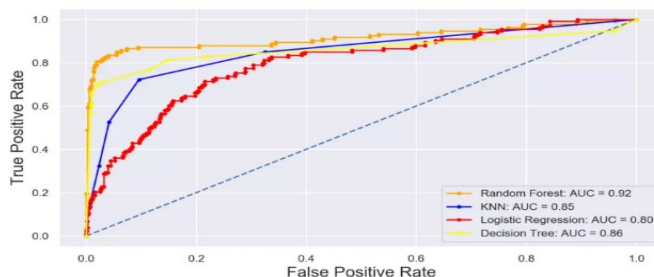
Резултати от работата на инструмента Confusion matrix, са дадени на фиг. 12, за сравнение на ефективността на създадените модели Tree и Logistic Regression.

Confusion matrix for Tree (showing number of instances)							
		Predicted					
		Bad	Excellent	Good	Middle	Very Good	Σ
Actual	Bad	13	0	1	1	0	15
	Excellent	0	19	0	0	2	21
	Good	0	1	12	0	2	15
	Middle	0	0	2	6	0	8
	Very Good	0	2	0	0	15	17
	Σ	13	22	15	7	19	76

Confusion matrix for Logistic Regression (showing number of instances)							
		Predicted					
		Bad	Excellent	Good	Middle	Very Good	Σ
Actual	Bad	14	0	1	0	0	15
	Excellent	0	20	0	0	1	21
	Good	0	0	10	1	4	15
	Middle	1	0	4	2	1	8
	Very Good	0	0	4	0	13	17
	Σ	15	20	19	3	19	76

Фигура 12. Резултати от работата на инструмента Confusion matrix за създадени два модели

Receiver Operating Characteristic (ROC) кривата съпоставя TPR (True Positive Rate) и FPR (False Positive Rate), което дава информация до каква степен моделът правилно разпознава съответните класове. На фиг. 13 е даден резултатът от съпоставяне на четири различни класификатори. Колкото по-близо до горния ляв ъгъл е ROC кривата, толкова по-високо е качеството на класификатора. На графиката се вижда, че в конкретния случай за предвиждане, моделът, използващ алгоритъм Random Forest се справя най-добре.



Фигура 13. Визуализация с инструмента ROC Analysis в Orange

Като се обобщят резултатите за средната оценка за прогнозната точност върху класовете оценки (Слаб, Среден, Добър, Много добър и Отличен) се получава: Decision Tree - 85.5%; Logistic Regression - 77.6%; Random Forest - 89.5%; Support Vector Machine - 84.2%; Naive Bayes - 81.6%; Neural Network 80.3.1%. Като подходящи за използване, при подобни задачи за оценяване, могат да бъдат класификационните алгоритми Decision Tree и Random Forest, които предвиждат с висока степен на точност елементите на Bad класа. Това са студентите със слаби оценки, които разглеждаме като „студенти в риск“.

В проведения експеримент инструментът обработва данните, които се събират по време на обучението на студентите в техните курсове. Освен това се използват и данни от анкетно проучване. Анкетата се изпраща по имейл през третата седмица от курса. В основата на анкетата са въпроси, с които студентите дават мнение относно задачите, материалите и нивото на трудност на учебния предмет. Данните, получени от анкетите, добавят редица нови характеристики, които са пряко свързани с отпадането на обучаемите, като липса на интерес, липса на време, организационни пречки по време на обучението и др. Първите експерименти са направени от малка извадка от данни, ограничена от първоначално налични реални данни за студенти. С помощта на разработената система за наблюдение и събиране на данни за обучаемите и тяхната дейност (решени задачи, преглеждане на електронни текстови и видео материали, участие във форум и/или групи по интереси и др.) от 4580 записа за обучаеми, 15 резултата от обучаемите са идентифицирани в риск. Това е многостъпкова процедура, която е пряко свързана с Big Data Analytics [101] в пространството за електронно обучение. Ежедневно се събира огромно количество данни, което сега изисква нов подход към обработката на данни, прилагайки рамка за разпределена обработка с отворен код Hadoop, базирана на алгоритъм Map/Reduce.

ИЗВОД: Проведените емпирични изследвания потвърждават, че предложеният **петстъпков метод за оценяване е перспективен** за разработване на система за ранно предупреждение за различни заинтересовани страни от обучаващия процес.

Базирайки се на натрупани данни от работата на система за електронно обучение с различни потребители, прилагайки средства от областта на науката за данните, могат да се взимат различни решения относно обучението. Анализът на данните, натрупвани в курсовете за електронно обучение дава възможност да се промени модела на изпитване и да се проектират модули, които да отговарят на индивидуалните нужди на обучаемия.

Подсилващо обучение (обучение с утвърждение)

Към алгоритмите за машинно обучение принадлежи и подсилващото обучение (*Reinforcement Learning*). Характерното за този тип алгоритми е, че имитира психологически модел, при който на системата се подават „награждаващи” и „наказващи” сигнали с цел да се максимизира вероятността за получаване на „награда” и да се минимизират „наказанията” [88]. Това обучение се прилага обикновено при липса на предварително зададен „правилен“ набор от обучителни данни. Такъв подход е различен от начина на действие при контролирано обучение, където целта е намаляване на отклонението, съобразно предварително зададени правилни данни (вход/изход).

Неконтролирано обучение

Алгоритмите от този тип приемат съвкупност от данни, съдържаща само входящи стойности и намират структура или разпределение на данните, без указания за известна променлива или функция за възнаграждане, данните не са етикетирани, няма примери за обучение [100]. Основни типове решавани задачи са: Dimensionality reduction, Density estimation, Clustering.

Клъстерният анализ представлява разпределяне на съвкупност от наблюдения в подмножества (clusters), така че наблюденията в един и същ клъстер да са сходни според един или няколко предварително зададени критерии, а наблюденията от различни клъстери да са различни [46]. Различните техники за клъстериране работят с различни предположения за структурата на данните, които често се дефинират чрез някаква метрика за сходство и се оценяват, например, по вътрешна компактност, или близост между членовете на един и същ клъстер, и разлика между клъстерите.

Разработените методи за групиране използват различен индукционен принцип. Farley and Raftery (1998) [32] предлагат разделянето на методите на групиране на две основни групи: йерархични и разделящи методи. Han and Kamber (2001) [45] предлагат категоризиране на методите в допълнителни три основни категории: базирани на плътност методи, базирани на модели и мрежови методи. Алтернативна категоризация, основана на принципа на индукция на различни методи за групиране е представена в Estivill-Castro (2000) [29].

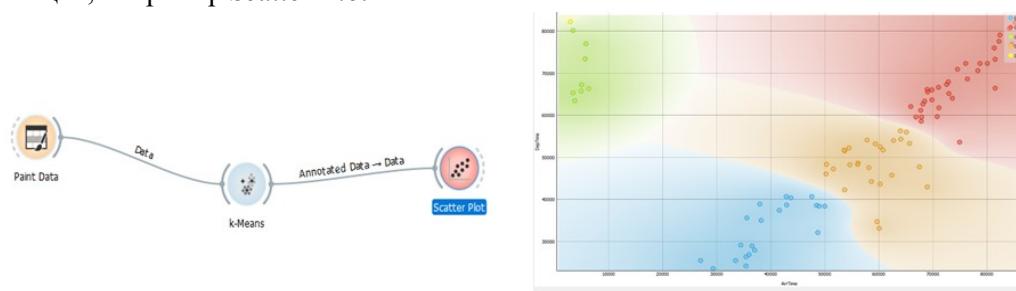
- Алгоритми за разделяне на клъстери

Тези алгоритми генерират различни дялове и след това ги оценяват по някои критерии. Те са наричани нейерархични, тъй като всеки екземпляр е поставен в точно един от k взаимно изключващи се клъстери. Изисква се предварително да се въведе желания брой клъстери.

Един от най-често използваните алгоритми за разделяне на клъстери е *k-means clustering* алгоритми. Този тип алгоритми принадлежат към *Exclusive Clustering*, защото данните са групирани, така че ако определена данни принадлежи към даден клъстер, тя не може да бъде включена в друг клъстер. Друг тип *Overlapping Clustering*, използва размити множества за групиране на данни, така че всяка точка може да принадлежи на два или повече клъстера с различна степен на членство.

Тристъпков алгоритъм: определя централната координата (центроид); определя разстоянието от всеки обект до центъра; групира обектите на базата на най-малкото разстояние. Прекратява се, когато центроидите спрат да се движат или някакъв праг е достигнат (например брой итерации). Следва примерен работен поток за прилагане на инструмент *k-means* върху набор от данни и визуализиране на резултата чрез инструмент *Scatter Plot* (фиг. 14).

За изграждане на информационния поток първо се извършва **събиране на данни**. Зарежда се работния плот и набора с инструменти и чрез инструмент „File” се зареждат данните за анализ. Може да бъде приложен и инструмент *Impute*” за почистване на данните. Следва **Избор на инструмент *k-means***. Този инструмент на системата *Orange* реализира *k-means clustering* алгоритъм. Последната стъпка е **визуализация на резултата**. Самият инструмент *k-means* не визуализира резултат, за тази цел трябва към информационния поток да бъде свързан инструмент за визуализация, например *Scatter Plot*.



Фигура 14. Работен поток с прилагане на инструменти *k-means* и *Scatter Plot*

- Йерархичната клъстериация

Когато броя на клъстерите не е предварително определен, се използват йерархични клъстеризационни процедури (*Hierarchical Cluster*). Тези алгоритми започват от деклариране на всяка точка за свой собствен клъстер и след това обединяват два най-подобни клъстера, докато не бъде удовлетворен критерий за спиране.

Голямото разнообразие на процедурите се поражда от използваната **метрика** между различните обекти. Йерархичната клъстериация изчислява йерархично групиране на произволни типове обекти от матрицата на разстоянията между тях и показва съответната *дендрограма*.

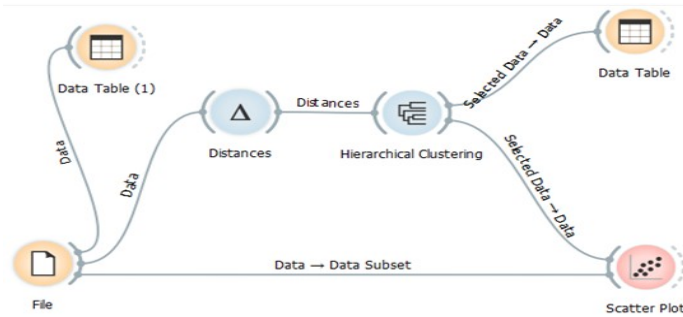
Метод за изграждане на информационен поток при йерархична клъстеризация е свързан с прилагане на следните *стъпки*:

1. **Събиране на данни.** Зарежда се работния плот и набора с инструменти и се зареждат данните за анализ. Може да бъде приложен и инструмент за почистване на данните.

2. **Избор на инструмент *Distance*** за изчисляване на разстоянието между елементите. За групиране на данните могат да се използват различни геометрични показатели за изчисляване на разстояние. *Euclidean metric* е мярка за разстоянието между точките, изобразени на евклидовата равнина. *Manhattan metric* е мярка, където разстоянието се изчислява като сума от абсолютната стойност на разликите между две точки, изобразени на декартовата координатна система. *Minkowski distance metric* е обобщение на показателите от *Euclidean metric* и *Manhattan metric*.

3. **Избор на инструмент *Hierarchical Clustering*.** Този инструмент реализира алгоритъм за йерархичното групиране на произволни типове обекти чрез изчислените разстояния и показва съответната дендрограма. *Дендрограмата* е граф-дърво, в което всеки възел отразява една стъпка от процеса на клъстеризиране.

4. **Визуализация на резултата.** Чрез различни инструменти може да се визуализира резултата по различни начини, например чрез инструментите *Data Table* и *Scatter Plot*.



Фигура 15. Работен поток за йерархично клъстеризиране

Други популярни софтуерни средствата за извличане на знания, които могат да бъдат използвани за анализ на данни са например: WEKA [103], RapidMiner [87]. KNIME [52], KEEL (<http://sci2s.ugr.es/keel/>), SPSS <http://www.ibm.com/analytics/us/en/technology/spss/> и др.

3.3. Уеб метрики за оценяване дейностите на обучаеми във виртуално образователно пространство

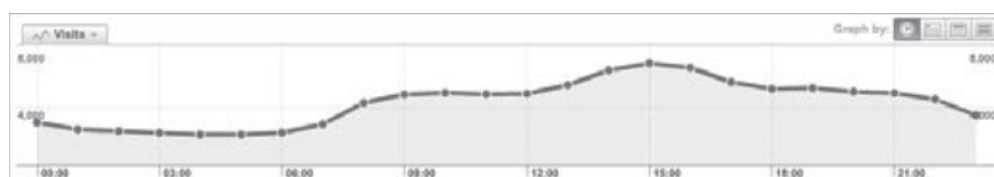
Анализът на данните, натрупани по време на обучението позволява проучвания и обратна връзка за това как обучаемите търсят информация, какви затруднения срещат и проектиране на модули, които отговарят на индивидуалните нужди на обучавания. На базата на веб метрики тук се предлага метод за оценка на степента на използване на веб сайтове, свързани с областта на обучението при моделиране на обучаемите, базирайки се на анализ на поведението им в обучаващата среда и веб пространството. Анализът на данните дава възможност да се проектират модули, които да отговарят на индивидуалните нужди на обучаемия за търсене на информация.

Уеб анализът е измерване, събиране, анализ и отчитане на интернет данни с цел по-добро разбиране на сложните взаимодействия между потребителите на уебсайтове. Процесът на уеб анализ включва различни уеб метрики, дефинирани в Google Analytics [25] като:

- **page views** - броя показвания на уеб страницата, достъпна от посетител (без паяци или роботи);
- **visitors** - броя на уникалните посетители на уеб сайта;
- **pages/visits** - броя на страниците, които са разгледани от посетител по време на посещение;
- **time on site** - продължителност на времето, прекарано от всички посетители на уеб сайта;
- **stickiness** - способността на уеб страницата да задържи посетителя на уеб сайта;
- **frequency** – (честота) брой посещения, от посетител на сайта (индикатор за лоялност);
- **recency** - брой дни, изминали от последното посещение на посетителя на сайта;
- **length of visit** - времето на посещение, прекарано от посетител на уебсайта (в секунди);
- **depth of visit** - брой страници, посетени от посетител за едно посещение и др.

В областта на уеб анализите се използват главно две техники за анализ на трафика на уеб сайта: събиране на данни от страна на сървъра и от клиентска страна. Методите за събиране на данни от сървъра извличат и анализират данни основно от регистрационните файлове (*Log files*) и включват информация за регистрационни файлове като IP адрес, час и дата, тип браузър и др. Информацията за поведението включва обща информация за сърфиране, като например брой прегледани страници, езикова настройка и др. Когато се събират данни от клиентски сайтове или страници с отметки, данните за посетителите на страницата се изпращат до проследяващ сървър с помощта на JavaScript код (или етикет), вмъкнат в HTML страницата. С този подход могат да се проследяват точно всички действия на посетителя, както и да се събира допълнителна информация. Бисквитите могат да се използват, за да се определи колко посетители за първи път или колко повторни посетители е получил даден сайт, колко пъти посетител се връща за всеки период и колко време минава между посещенията.

Google, WebTrends, Nedstat [5] и много други компании предоставят софтуер за уеб анализ с помощта на маркиране на страници. Google Analytics е най-използваната безплатна програма [25]. От гледна точка на анализа на дейността на обучаемите, особено интересна е свързаната с клиентите информация, историята на посещенията и поведението им, потребителския профил и др. Google Analytics изготвя анонимни и статистически отчети за уебсайтовете, които го използват. Показват данни като географско местоположение (на базата на общи кодове за геолокация, базирани на IP), време на посещение и т.н. Например: отчетът на фиг. 16 [25] дава възможност да се проследи по часове посетителския трафик на сайта.



Фигура 16. График за почасова тенденция с Google Analytics.

За да се направи анализ на най-посещаваните от обучаемия уеб сайтове, дълбочината и честотата на посещенията, се използва *Fuzzy Classification of the Web Metric* [107]. Известно е, че в теорията за размитите множества и размитата логика, се отчитат неточност, несигурност и неяснотата на човешкото мислене и език, определяйки функцията на принадлежност.

Размито множество A в X се дефинира като [105]:

$$(9) \quad A = \{(x, \mu_A(x))\}, \text{ където } x \in X, \mu_A : X \rightarrow [0, 1]$$

е функцията на принадлежност на A и $\mu_A(x) \in [0, 1]$ е степента на принадлежност на елемента x в множество A . В нашите експерименти за броя на посещенията на даден уеб сайт, се въвеждат термините „low“, „medium“ и „high“ като езикови променливи. В общия случай, отчитайки стойностите за показване на страниците като true (1) или false (0) за месец, страниците се определят като „ниско посещавани“, ако посещенията са например между 0 и 25, между 25 до 55 посещения страниците са „средно посещавани“ и повече от 56 посещения са класифицирани като „високо показвани“. Въпреки това, ако посетител има 55 посещения на страница, той се определя в клас „средно“ активен, докато друг посетител с 58 посещения се определя като „много“ активен посетител. Въпреки разлика само от само от 3 посещения, те са разпределени в две различни групи. Чрез дефиниране на размити множества, представени с функция за принадлежност се въвежда непрекъснат преход между класовете „low“, „medium“ и „high“. Така посетител може да принадлежи отчасти на два класа (55% на „високо активен“ и 45% на „средно активен“) едновременно. Използването на размити класове позволява по-прецизна класификация на уеб метричните стойности [50]. Отчитайки това с цел а да се оцени трафикът на уебсайта, се разработва размита система от правила [28]. Измерването на броя на посещения на страница е число, което само по себе си няма голямо значение. Само контекстът на числото спрямо броя на прегледите на други страници предоставя знание, на базата на което можем да направим оценка.

За определяне на броя на посещенията на уеб страници се реализират правила за размита класификация. Прилага се метода на *индуктивна размита класификация (Inductive Fuzzy Classification IFC)*, при който групирането на елементите в размит набор, се извършва с функция за членство, извеждана чрез индукция от данни. Индуктивната размита класификация по процентилен ранг (IFC-PR) [50] генерира размита функция на членство, в случая използвайки общите езикови термини „low“, „medium“ и „high“, корелиращи с броя потребителски посещения:

- емпиричният ранг на стойността x с метриката M определя принадлежността към размития клас „high“: $\mu_{\text{high}}(x) = P(M < x)$

- показателят M ще бъде класифициран като „low“ (отрицанието на принадлежността към класа „high“): $\mu_{\text{low}}(x) = 1 - \mu_{\text{high}}(x)$

- класификация „medium“ се дефинира като:

$$\mu_{\text{medium}}(x) = 1 - \text{abs}(\mu_{\text{high}}(x) - 0.5) - \text{abs}(\mu_{\text{low}}(x) - 0.5).$$

В провежданото изследване [82] се прави анализ на посещенията на уеб сайтовете, указани като помощни материали в курса по дисциплината Изкуствен интелект. За всяка уеб страницата се оценява посещаемостта, като например: страницата $W1$ има 135 посещения в рамките на обучаващия курс. Общо предложените и наблюдавани страници в курса са 80 и 56 от тях имат по-малко посещения от $W1$. За оценка на посещенията на $W1$, се пресмята:

$$\mu_{\text{high}}(\text{посещение } (W1)) =$$

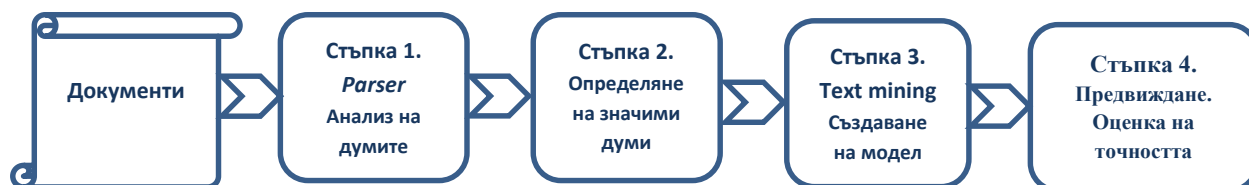
$$P(\text{Брой посещения} < \text{посещения } (W1)) = 56/80 = 0,7.$$

Така според размитата класификация се определя, посещаемост на страницата „висока“ над 0,7, за „ниска“ под 0,3, а за „средна“ до 0,6. На базата на получения анализ за всяка уеб страница, предлагана в обучаващия курс, се взимат решения за актуализиране на уеб източниците,

както и допълване на литературата с нови сайтове, свързани с най-търсените теми.

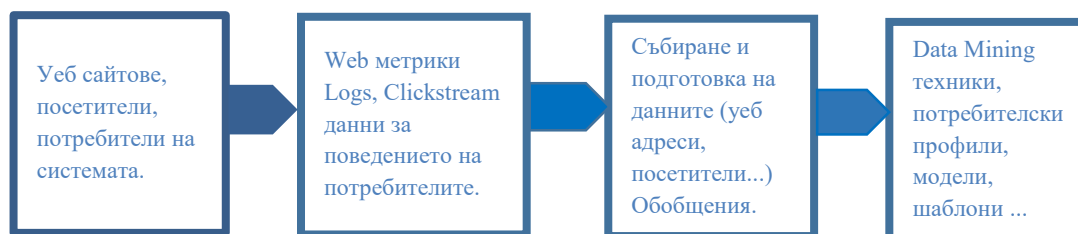
След като се направи анализ на най-посещаваните от обучаемия уеб-сайтове, дълбочината и честотата на посещенията, следва да се анализира и определи типа на тяхното съдържание. Подход за автоматизирано определяне типа на документ и доколко е свързан с областта на обучението е представен в [80]. Анализът се базира на идеята: определени думи в документа са значими за съдържанието му. За да се определят тези думи се търси честотата на появата им, използвайки софтуера *Wordcloud* [104]. Могат се налагат филтри, относно получавания списък с думи. Изчислява се близост на всяка дума от полученото множество, с всички думи от речника на предметната онтология, свързана с областта на обучение и се взима най-малката получена стойност за всяка дума. За определяне на близост на думите и фразите се използва подход, представен в [48] чрез q-gram метрики.

Чрез обработка на голямо множество от уеб-документи се натрупва набор от данни с извлечени от тях централни думи и степента им на близост с понятията от анализираната област. Тези данни са използвани за обучение на класификационни алгоритми, представени и описани под формата на многостъпков процес, като целевия атрибут е дали документа е свързан с областта или не. Предложен е метод за анализ на документи, показан с схемата на фиг. 17.



Фигура 17. Многостъпков процес на анализ и прогнозиране на документи от зададена област

От събираната информация за дейността на потребителите в уеб пространството с помощта на клъстерен анализ или прилагане на асоциативен анализ, може да се извършва потребителско сегментиране. Уеб съдържанието може да бъде анализирано на следните стъпки (фиг. 18).



Фигура 18. Архитектура за извличане на данни от уеб метрики

ИЗВОД: В етапа на анализ на резултатите от проведеното обучение, особено внимание се отделя на трудностите, които обучаемите срещат при изпълнение на отделните модули в дисциплината. В проведения експеримент, направените изводи са свързани с отчитане на факта, че студентите в изследваната дисциплина Изкуствен интелект срещат трудности при работа с литературните източници за изготвяне на проекта по избрана тема от *Модул 4*, в областите: „Blockchain“, „Ethics and Emerging Sciences“, „Policy and investment recommendations for trustworthy Artificial Intelligence“ както и при определяне на рискове, свързани с тези технологии.

Чрез представения метод в средата за обучение може да се проследи активността на обучаемите в интернет пространството, като се натрупват данни, относно тяхната активност в обучаващата среда. Чрез анализ на този тип данни се получава навременна информация за активността, напредъка и успеха на обучаемите. Резултатите показват, че формулираните решения могат успешно да се използват за различни задачи и могат да бъдат адаптирани към нови технологии и приложения. Предложеният метод насърчава създаването на иновативна обучаваща среда и е следващата стъпка в дигитализацията на образованието.

3.4. Моделиране на потребителя в образователното пространство

Чрез подходящи средства за следене и отчитане на дейността на потребителя се изгражда негов модел. Моделът служи за персонализиране на системата към знанията и уменията на потребителя и прилагане на адекватни подпомагащи стратегии в работата му. **Моделът на потребителя** по своята същност е структура от данни, представяща индивидуални характеристики за потребителя, на дадена програмна система. Той представя познавателните процеси и схващания на потребителя за приложната област [7].

Според Barr и Feigenbaum [10] съществуват два **базови подхода**:

- При *моделите с припокриване* (overlay approach) знанието на потребителя се представя като подмножество на общото знание, поддържано от системата за областта, която се изучава, знанието на експерта в областта или очакваното знание за обучаемия. Знанието на системата е декомпозирано на независими компоненти и е покрито със система от означения, показващи нивото на усвояване на всяка отделна компонента. Приемайки това становище тук се изгражда модел за оценяване на базовите знания на обучаемите относно основните понятия и зависимости между тях. В преподаваната предметна област е изградена йерархия на понятията, дефинирана в използваната домейн-онтология. След всеки тест на понятията се приписва относителна числова стойност (в проценти), която показва степента на сигурност на системата относно познаването на това понятие от конкретния обучаем. Оценката за всяко понятие може да се формира динамично като осреднена оценка на дъщерните му понятия по формулата:

$$(10) \quad \text{Mark_term} = \frac{1}{k} \sum_{i=1}^k \text{Mark_Subterm}[i],$$

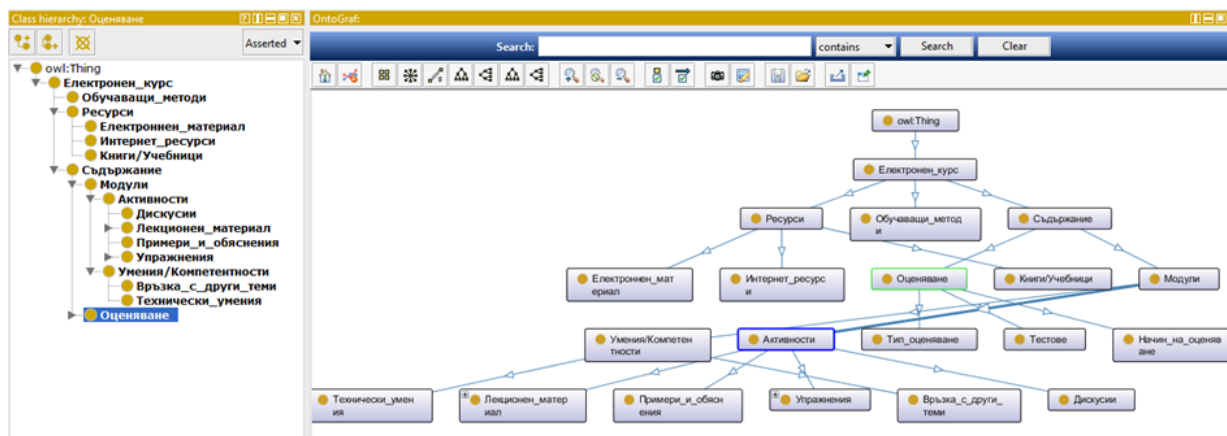
където $\text{Mark_Subterm}[i]$ е оценката на i -тото дъщерно понятие, а k е броят на дъщерните (наследените) понятия на оценяваното основно понятие. Всяко дъщерно понятие от своя страна може да се разглежда като родителско на своите дъщерни понятия и да получава оценката си по същата формула. Тогава общата оценка за обучаваната тема може да се формира по формулата:

$$(11) \quad \text{Mark} = \frac{1}{n} \sum_{j=1}^n \text{Mark_Term}[j],$$

където $\text{Mark_Term}[j]$ е оценката на j -тото основно понятие в изучаваната тема от предметната област, а n е броя на тези основни понятия.

- При подхода *моделиране с библиотека от грешки* (bugs theory) формално се диагностицират грешки, чрез списък от предварително дефинирани грешно усвоени и липсващи елементи от знания. Тук моделът на потребителя се състои от модел на експерта, допълнен със списък от допускани грешки в областта. Голяма част от тези данни се събират от средата за електронно обучение и се съхраняват в реляционна база от данни. Използвайки натрупаните данни, се генерират справки за участниците и работата им, например: справка за участниците в дадена дейност, справка за продължителност на дадено обучение, дата на започване и дата на завършване на участието; данни за обучаемите и курсовете, в които участват; общ брой участници; брой участници, преминали обучение в отделните курсове и др.

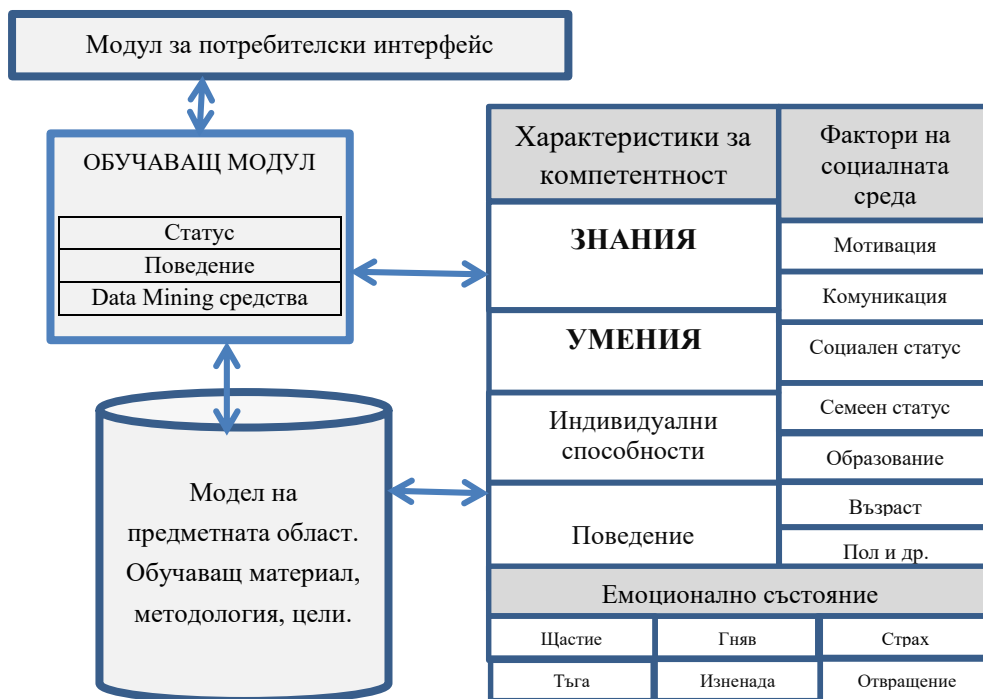
За да може обучението да бъде формално представено на мета-ниво, то неговите базови аспекти (знание, процеси и участници) трябва да бъдат моделирани и формализирани по начин, който осигурява дефиниране, намиране и използване на релации между тях. В дисертационния труд се разработва и предлага модел на учебния процес в образователното пространство, базиран на онтологии за представяне на знанията. На фигура 19 е представена схема на разработвания онтологичен модел на електронен курс. Прототип на онтологията е тествана чрез системата Protege (<http://protege.stanford.edu/>).



Фиг. 19. Онтологичен модел на електронен курс чрез системата Protege

Моделът на знанията от областта на обучение представя основните концепции и техните връзки, които трябва да бъдат преподавани и проследявани като ниво на усвоеност.

На базата на експерименти с обучаеми в електронна среда по различни дисциплини тук се предлага обща рамка на архитектура на модел на обучаемия, който включва три типа фактори: фактори на компетентност, емоционални фактори, фактори на въздействие на социалната среда.



Фиг. 20. Архитектура на обучаваща система с модел на обучаемия

Предложеният модел на потребителя е обособен като отделен модул, съдържащ данни за потребителя, и система от функции, грижещи се за събирането, съхранението, обработката и интерпретацията на тези данни. Този модул е свързан с модела на предметната област на обучение и обучаващият модул на системата. В индивидуалния модел (фиг. 20) се съдържат три типа данни.

- Първият тип представя оценки относно елементарни умения и типични грешки при работа в областта на обучение. Тези оценки се обобщават в общата оценка на обучаемия в областта.

- Вторият тип показва нивото на усвояване от обучаемия на всяка от единиците знание (ЕЗ). За обучаем се поддържа списък от елементи от вида: <номер на ЕЗ> : <оценка на ЕЗ>.

- Третият тип касае работата на обучаемия по време на обучаващата сесия: съдържа номера на учебна програма, текуща трудност на въпросите и задачите; нивото при оказване на помощ и др.

Тук се включват и данни за обучаемия като: пол, възраст, културен и социален статус (включва публикации във форум и брой отговори, групово взаимодействие), както и някои индивидуални особености. Тази компонента е в процес на експериментални изследвания и затрудненията произтичат от факта, че не е известна количествена теория за връзките между психологическите характеристики на обучаемия.

Оценките указващи нивото на усвоеност или неувоеност на дадено знание или оценките на въпроси, задачи или заявки на обучаемия, представляват реални числа в интервала $[-1,1]$.

Комбинирайки този подход с подхода за поддържане на библиотека от грешки, за всеки от предварително дефинираните типове грешки се поддържат броячи за случаите, когато е било възможно да бъде допусната грешка и броя на грешните опити. След приключване на събитието за всеки от типовете грешка се формира оценка по формулата:

$$(12) \quad Mark_{current} = 1 - Count_error / Count$$

Тук $Mark_{current}$ е междинна оценка, която е число в интервала $[0,1]$ и всъщност представлява процента на опитите от страна на потребителя, при които не е допусната грешка. $Count$ е общият брой опити, а $Count_error$ е броят на грешните опити. За да се получи окончателна оценка $Mark$ за уменията в интервала $[-1,1]$, се пресмята по формулата:

$$(13) \quad Mark = 2 \cdot Mark_{current} - 1.$$

Вземайки предвид предните две формули, за оценката получаваме:

$$(14) \quad Mark = 1 - 2 \cdot Count_error / Count.$$

В случая, когато общият брой опити е 0, на оценката присвояваме стойност 0. След като е формирана оценка на типа грешка след събитието, тя се отразява на оценката за съответния тип грешка в модела на потребителя по формулата:

$$(15) \quad Mark_{new} = (1 - K) \cdot Mark + K \cdot \text{Оценка_на_задачата}.$$

Тук K е число в интервала $[0,1]$ и представлява влиянието, което оценката на типа грешка указва върху оценката в модела на обучаемия. Пресмята се по формулата:

$$(16) \quad K = 0.5 \cdot (Count - 1) / Count$$

Така в оценката за всяко от уменията се съдържа информация от предишни задачи и от последната задача. Оценката се умножава по 0.5, за да се ограничи влиянието на старата оценка.

Пресмятането на оценките от втория тип за нивото на усвоеност на отделните единици се извършва като, за всяка решавана задача или поставен въпрос, оценката се променя по формулата:

$$(17) \quad \text{Оценка_нова} = (1 - K) \cdot \text{Оценка_стара} + K \cdot \text{Оценка_на_проблема}$$

Но тук $\text{оценка_на_проблема}$ е оценката на въпроса или задачата, задаена от преподавателя. K е влиянието, което оценката на проблема оказва върху оценката на единицата знание и се пресмята по формулата:

$$(18) \quad K = \text{Трудност_на_проблема} / \text{Трудност_на_ЕЗ}$$

колкото трудността на проблема (въпрос/задачата) е по-голяма, толкова по-голямо влияние оказва върху оценката на единицата знание.

Оценките са в интервала $[-1,1]$, когато оценката е по-близо до 1, то системата е по-сигурна, че обучаемият е усвоил вярно дадената единица знание. Колкото оценката е по-близо до -1, то системата е по-сигурна, че обучаемият е усвоил грешно дадената единица знание. Когато оценката е близо до нула, системата не е сигурна в знанията на обучаемия. Така се отчитат характерни черти и на подхода “моделиране с припокриване” и на подхода “моделиране с библиотека от грешки”, защото: ако оценката е положителна, тя представя нивото на знанията на обучаемия в сравнение със знанията на експерта, ако оценката е отрицателна, тя представя грешно усвоени знания.

Използват се броячи, с които се следи броя на използваните обучителни ресурси и типа на посетените интернет ресурси, които в различни комбинации могат да осигурят различно представяне на учебните материали. Информационните ресурси се оценяват по брой, видове и др.

Разпознаване на емоционално състояние на обучаемия в среда за електронно обучение

За целите на провежданото изследване в дисертационния труд е разработено софтуерно приложение за разпознаване на човешки емоции по образа на лицето. Разработеният софтуер се базира на системата за кодиране на лицевото действие (FACS) на Ekman и Friesen [27]. Лицевият израз на емоциите се определя от включените мускули и жестовете, които го характеризират. Съществуват определени отличителни модели на емоционални реакции, обобщени и споделени от повечето хора. Те се считат за основни емоции: щастие (*Joy*), тъга (*Sadness*), гняв (*Anger*), изненада (*Surprise*), страх (*Fear*) и отвращение (*Disgust*). Други подобни кодиращи системи са EMFACS, MAX, AFFEX и CANDIDE-3. Разработеният софтуер и проведеното изследване са представени в [57].

Методология. Първата стъпка е категоризиране на зададените снимки със изображения на лица, в отделни групи (поддиректории). Те са седем на брой (включвайки и *Neutral*), като всяка за от тях отговаря за конкретна емоция и съдържа около петдесет примерни изображения, свързани със съответната емоция. Следваща стъпка е използването на *Dlib* библиотеката от алгоритми за създаване на софтуер с общо предназначение, която се използва в много области. Чрез използване на файла *shape_predictor_68_face_landmarks.dat*, лицата от изображенията се маркират със съответни основни точки. На базата на поставените точки, в следващата стъпка се извършват математически изчисления, с помощта на които се определят 6 нормализирани вектора на всяко лице. Всеки вектор с лицеви характеристики е изчислен като данните се нормализират и се съберат разстоянията между маркираните точки. Чрез средствата на framework „ML.Net“ се имплементира многокласов класификационен алгоритъм, базиран на ML агент от тип Supervised Learning.

След обучението на модела, той може да разпознава емоционалното състояние на подадено изображение (снимка) на лице. В резултат на работата, програмата връща предвиждане относно емоционалното състояние. Анализът на резултатите от метриците за оценка на точността на модела показва, че най-голяма точност на предвиждането се постига при емоционалните състояния *Joy*, *Surprise* и *Sadness*. Състоянието *Anger* често се бърка от алгоритъма със *Sadness*, *Joy* или *Neutral*. Алгоритъмът може да бъде подобрен чрез обучение с много по-голям обем тренировъчни данни от различните типове емоционални състояния. Но трябва да се използва балансиран набор от данни.

Алтернативен подход е прилагане на алгоритъм от тип *Unsupervised Learning*, при който се изследва структурата на данните без указания, данните за обучение не са етикетирани. Могат да се използват на *Clustering* алгоритми или *Deep Learning*. Прилагайки различни класификатори може да се експериментира с получаваните резултати за повишаване на точността, спрямо различни групи и категории данни. Целта е разработването на приложение да се интегрира към среда за e-Learning и да се търсят промените в емоционалните състояния на учащите се, докато се справят с когнитивните задачи в процеса на обучение.

Устройство за подпомагане на потребители с увреден слух

Проектирано и експериментално изследвано е устройство [106], което произвежда широка гама от цветове и нива на интензитет, за да визуализира различни аспекти на звука. За хора с увреден слух, звуковата визуализация осигурява подобро усещане за заобикалящата среда и им позволява да интерпретират и да реагират на слухова информация в реално време.

Разработени са софтуерни инструменти за анализ на звуковите честоти и преобразуването им в съответните цветове в RGB модела. Създаден е модел на автоматизирана система, която реализира преобразуване на звуковите честоти в цвят. Постигнатата точност е оценена експериментално. Установено е, че точността на преобразуването зависи от честотата на звука. Доказано е, че процентът на грешки е по-висок при високи честоти на звуковите вълни (над 1000 Hz). Работа трябва да продължи с изследвания, насочени към обобщаване на моделите за преобразуване на звуковите честоти в цвят.

3.5. Когнитивен агент в среда за електронно обучение

Когнитивният агент [24] е автономен софтуер, който има способността да развива своите знания. Като се има предвид различната степен на интелигентност, когнитивните агенти могат да бъдат реализирани на базата на правила, които развиват знанията си с помощта на вградени в системата механизми за извод или извличат нови знания с помощта на техники на машинно обучение и Data Mining техники.

Общата рамка на архитектурата, на изгражданата за това изследване система, базирана на правила - *Rule Based Systems* включва компонентите [89]:

- о *База от знания* (множество от правила) - съдържа всички знания необходими на агента.
- о *База от данни*: съдържа данните, които са установени към текущия момент.
- о *Машината за извод (Inference Engine)* - интерпретаторът избира и прилага правилата, които могат да доведат до промяна на състоянието на базата от данни. Съществуват различни подходи: свързване напред, свързване назад и смесени подходи [89].

о *Базата от мета-знания* - мета-правила, задаващи връзката между правилата [24]. Начинът, по който се избират правилата от множеството на приложимите правила е стратегията за решаване на конфликти (*Conflict Resolution*). В случая се прилагат приоритети на правилата, които са динамични. Динамичните приоритети се определят въз основа на важността на действията в настоящата ситуация. Например: ако *оценката на обучаемия е висока*, то се задава нисък приоритет на правилото за търсене на следваща задача за решаване. Тогава обучаемият ще търси следваща задача за решаване само, ако няма други дейности за изпълнение. Но ако е със *слаба оценка*, правилото за решаване на следваща задача трябва да има висок приоритет. Обучаемият трябва да решава следваща задача, за да повиши оценката си.

От друга страна обучението с утвърждение (*Reinforcement Learning*) става все по-популярен тип машинно обучение. Това е техника, при която автономните агенти използват алгоритмите за проба-грешка и функция за кумулативно награждаване на обучението. Предимството се състои в калкулирането на оптимални действия, които агентите могат да предприемат в рамките на сценарии, обусловени от средата. Такъв подход е коренно различен от начина на действие при контролирано обучение, където целта е намаляване на отклонението съобразно предварително зададени правилни данни (вход/изход). Обучение протича като последователност от пробни действия, които постепенно водят до утвърждаване на добрите действия и избягване на неподходящите. Резултатът от обучението е оптимална стратегия за действие във всяка ситуация. Стратегията е оптимална, ако успява да максимизира сумата от всички награди, получени по време на изпълнението си.

Формално обучението чрез *Reinforcement Learning* се разглежда като [54]:

- Множество S , включващо всички състояния на обкръжението, които агентът разпознава.
- Множество A , включващо всички действия, които агентът може да извърши.
- Множество от правила за преходи между състоянията.
- Правила, които определят R - наградите, която агентът може да получи при преходите.
- Правила, които описват какво спазва агента.

Конкретна задача в дисертационния труд е да се анализира и сравни поведението на два агента, реализирани на базата на двата различни подхода, базирани на машинно обучение. Резултатите от проведеното изследване и разработеният за целта софтуер са представени в [58]. За целите на анализа е създадена игра, която предоставя платформа и генерира x брой бомби и монети върху нея. Играта създава два агента, представени като две различни топки на платформата – червена и синя. Платформата няма оградящи стени, което означава, че топките (агентите) могат да паднат. Ако агент падне, той губи играта. Всеки агент може да събира монети, като всяка монета носи на агента точка. Ако агент докосне бомба, той губи точка. Играта

приключва, когато монетите свършат или когато агент падне от платформата. Агентът с повече точки печели играта. Един агент, представен със синя топка в играта, поддържа поведение въз основа на система от правила, създадени от програмиста и вградени в базата знания на системата. В този случай при изграждането на играта се прилага подходът *Rule-Based System*. От друга страна, агент, представен с червена топка, се реализира въз основа на подхода за машинно обучение с *Reinforcement Learning*. При него изкуственият интелект научава сам кое е най-доброто действие във всяка ситуация и с времето оптимизира решенията, които взема.

Разгледани са разликите в поведението на създадените решения чрез избраните два подхода. Хипотезата е, че агентът, реализиран с машинно обучение чрез *Reinforcement Learning* [54] ще има по-разнообразно поведение и след достатъчно време за обучение ще работи по-ефективно от агента, реализиран като система, базирана на правила.



Фигура 21. Изглед на разработения софтуер

Повече експерименти и разбиране на типа машинно обучение *Reinforcement Learning*, допринасят за разширяване на обхвата на използването му в съвременните среди за обучение. От друга страна алтернативата на системите, базирани на правила е подход доказал във времето своите предимства и ефективност за бързо постигане на поставените цели. Проектът [58] е създаден с помощта на платформата *Unity*, заедно с пакет на *Unity Technology* за разработване на агенти за машинно обучение *ML-Agents*. Този пакет предоставя сценарии за обучение и валидиране. Всеки агент има набор от състояния и наблюдения, може да предприема действия в средата и да получава награди.

- *Методология на изследването.*

Първият агент с *Reinforcement Learning*, е създаден чрез наследяване от клас *Agent*. Предефинира се методът *OnEpisodeBegin*, в който се задават първоначалните параметри за играта (монети и бомби и техните позиции). Предефиниране ба метода *CollectObservations*, осигурява предаване на текуща информация за посоката към най-близката монета. *OnActionReceived* метод задава разрешените движения на агента. Когато агент докосне обект, действието се засича и се извиква функцията "*OnTrigger*", която позволява да се види с кой обект агентът е влязъл в контакт.

Вторият агент, базиран на система от правила, е имплементиран чрез *RBS* класа и дейността на агента се задава от реализираните методи в класа. В зависимост от сценариите се прилагат правилата: „Ако има монети на платформата, отиди да вземеш най-близката“; „Ако RBS агента има повече монети от другия агент, той не трябва да се страхува от бомбите, за да завърши играта по-бързо и да не позволи на противника да събере повече монети. Изпълнението на правилата се основава на подхода свързване напред (*forward chaining*).

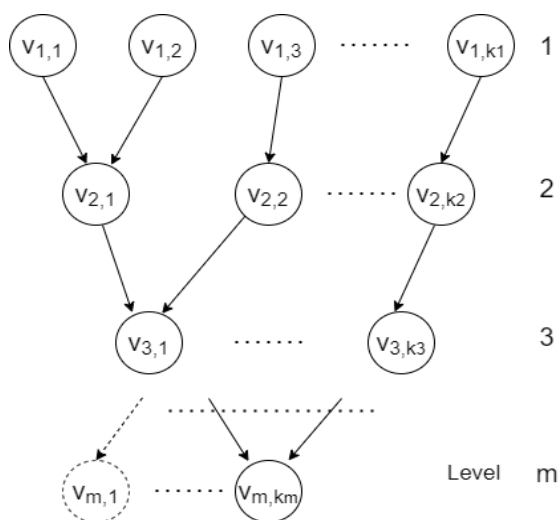
ИЗВОД: *Reinforcement Learning* агентът постепенно научава кое е най-доброто действие във всяка дадена ситуация и с времето оптимизира решенията, които взема. Той първоначално работи неефективно, но с течение на времето ще оптимизира действията си. От друга страна, при условие, че са създадени достатъчно добри правила, *Rule-Based System* е програмна задача, която може да поддържа много сложно поведение. Интересна тенденция в по-сложните приложения е съвместно прилагане на обучение с утвърждение с контролирано и неконтролирано обучение, ако използваните самостоятелно методи не дават достатъчно добър резултат.

3.6. Модели за йерархично многокомпонентно оценяване на обучаеми

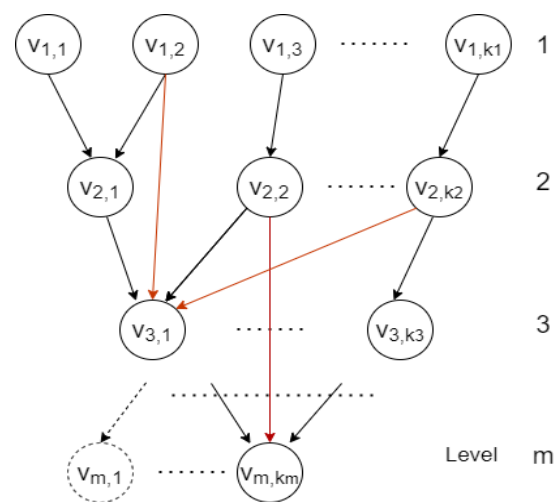
Представени са **нови модели** за йерархично многокомпонентно оценяване на обучаемите, които имат за цел комплексно оценяване на различни мисловни умения от висок и нисък порядък, теоретични знания и практически умения, и др.

Таксономията на Блум [13] дефинира йерархия на мисловните умения, в която по-високите нива на мислене включват всички познавателни умения от по-долните нива. Нивата са структурирани като: *Knowledge* (Знания), *Comprehension* (Разбиране), *Application* (Приложение), *Analysis* (Анализ), *Synthesis* (Синтез) и *Evaluation* (Оценяване). Всяко ниво се определя от множество когнитивни умения, дейности и методи за оценка. Счита се, че в процеса на обучение, обучаемият преминава последователно през всички нива. Една общоприета класификация определя уменията от горните три нива на Блум (*Analysis*, *Synthesis* и *Evaluation*) като *higher-order thinking skills* (HOTS), а тези от долните три нива (*Knowledge*, *Comprehension* и *Application*) – като *lower-order thinking skills* (LOTS). HOTS включва способности за критично мислене, умения за трансфер на знания и умения, умения за решаване на проблеми, и др. [15]. Оценката на студентите трябва да е обективна, а в много случаи е трудно, особено когато се отнася за оценяване на HOTS, които изискват креативно мислене. Това мотивира работа в насока създаване на комплексни многокомпонентни модели за оценяване на студентските постижения [36].

Предлага се **моделиране на йерархичната организация на оценъчните компоненти и зависимостите помежду им, които оценителят явно или неявно използва при оценяване**. Например, при множество дисциплини, основни оценъчни компоненти са „практика“ и „теория“, а в използваните термините тук: **thinking skills**, съответни техни под-компоненти могат да бъдат теоретични и практически **LOTS** и **HOTS**. От друга страна ако изберем за основни компоненти HOTS и LOTS, те биха могли да имат под-компоненти за оценяване на теоретични знания и практически умения. Създаването на конкретна йерархия от компоненти и подкомпоненти може да бъде извършено с помощта на предварително стандартизирани модели, избрани от оценителя. При достатъчно входно-изходни образци и ефективно обучение на невронна мрежа, получените резултати могат да отразяват тази йерархична организация, дори без тя да е указана явно.



Фигура 22. а) **Модел 1:** Йерархична дървовидна организация на оценъчните компоненти



Фигура 22. б) **Модел 2:** Йерархична графова организация на оценъчните компоненти

На фиг. 22 а) е представен общ **Модел 1 на йерархична дървовидна организация на оценъчните компоненти**, при който всеки основен компонент, в частност и крайната оценка, зависи от множество под-компоненти на предходното ниво. При това, всеки под-компонент влияе само върху един компонент от по-високо ниво.

Модел 2 на йерархична графова организация на оценъчните компоненти (фиг. 22 б), разширява и обобщава възможностите на Модел 1, с възможността един компонент да влияе на множество компоненти от произволни нива и да зависи от компоненти на различни по-ниски нива.

Представените модели могат да се използват при множество различни подходи за оценяване, при които крайната оценката е функция на множество оценъчни компоненти. Основни характеристики на предложените модели са:

- Ниво 1 на компонентната йерархия описва конкретни стойности от проведени оценявания, които могат да бъдат с различни оценъчни скали. Примери за такива стойности са: точки от тестове за теория и практика; точки от тестове за умения, оценки от задачи, проекти и др.

- На всяко следващо ниво, компонентите формират оценка, която е функция на оценките на подчинените му под-компоненти от предходното ниво. В частност, всички или част от функциите могат да бъдат с размитата логика.

- На последно ниво има един компонент, т.е. $k_m=1$, но не е задължително. Възможно е да има няколко крайни оценки, напр. за различни оценявани когнитивни умения. При това $k_m \geq 1$.

Възможен вариант за функции на ниво 2 е нормирането на всички стойности на компонентите от ниво 1 до една обща система за оценяване (например от 2 до 6). Аналогично, функциите на всяко следващо ниво биха могли да запазват нормираността на родителските компоненти в съответната скала за оценяване.

Обикновено, оценъчните под-компоненти имат стойности в пространството на реалните положителни числа R^+ . При различни подходи за оценяване, стойностите могат да бъдат и в други пространства на реални, комплексни или други числа. Крайната оценка E задължително принадлежи на пространство от предварително определени възможни стойности – например тя трябва да е цяло число в интервала от 2 до 6, от 1 до 5, от А до F, и др.

При формалното математическо описание на **Модел 1** и **Модел 2**, всяко ниво с номер i притежава k_i на брой компоненти, определящи множеството:

$$(19) \quad V_i = (v_{i,1}, v_{i,2}, \dots, v_{i,k_i}) \in R^{k_i}, k_i \in N, i = 1, \dots, m$$

Компонентите на ниво 1 са конкретни стойности, получени при оценяване. На всяко следващо ниво $i = 2 \dots m$ стойностите се получават в резултат на прилагането на функции върху компонентите от предходното ниво. В **Модел 1** броят на компонентите в дадено ниво намалява или се запазва равен на броя на компонентите в предходното ниво $k_s \leq k_{s-1}, s = 2, \dots, m$, което е следствие на обединяването на под-компонентите в компоненти. Броят на обобщаващите функции и съответно на компонентите, може да нараства в определени нива при **Модел 2**.

Стойностите на компонентите, след ниво 1, при **Модел 1** се изчисляват като **функция** върху стойности на компонентите от преките родителски нива:

$$(20) \quad v_{i,j} = f_{i,j}(P_{i-1,j}), P_{i-1,j} = (v_{i-1,s_1}, v_{i-1,s_2}, \dots, v_{i-1,s_j}) \subseteq V_{i-1}, \\ \forall i = 2, \dots, m, j = 1, \dots, k_i, s_j \leq k_{i-1} \text{ и}$$

компонентите на V_{i-1} се срещат само веднъж в някое $P_{i-1,j}$.

При **Модел 2**, параметри на функциите на дадено ниво, могат да бъдат произволни оценъчни компоненти от всички предходни нива, като при това отпадат множество ограничения:

$$(21) \quad v_{i,j} = f_{i,j}(Q), Q \subseteq \bigcup_{r=1}^{i-1} V_r, \forall i = 2 \dots m, j = 1 \dots k_i.$$

Експеримент с използване на *fuzzy logic* при йерархично многокомпонентно оценяване

В проведения експеримент по дисциплината „Програмиране в Интернет“, чрез четири изпитни компонента, оценяваме теоретични знания и практически умения за програмиране:

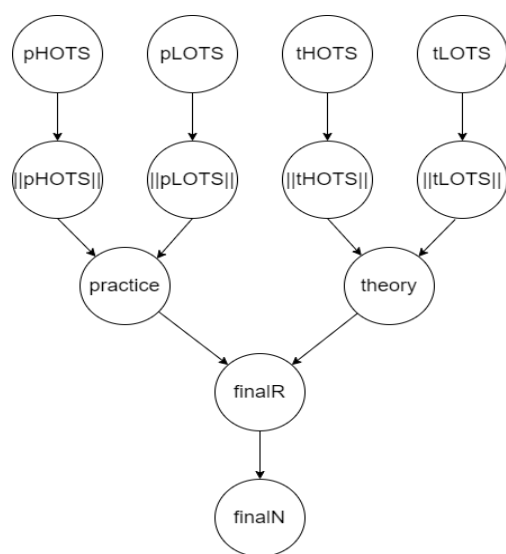
- tLOTS – теоретични LOTS, оценяват се чрез точки в интервал $[0, 30]$;
- pLOTS – практически LOTS, оценката е цяло число в интервала $[0, 15]$;

- tHOTS – теоретични HOTS, оценката е цяло число в интервала [0, 15];
- pHOTS – практически HOTS, оценката е реално число в интервала [2, 6].

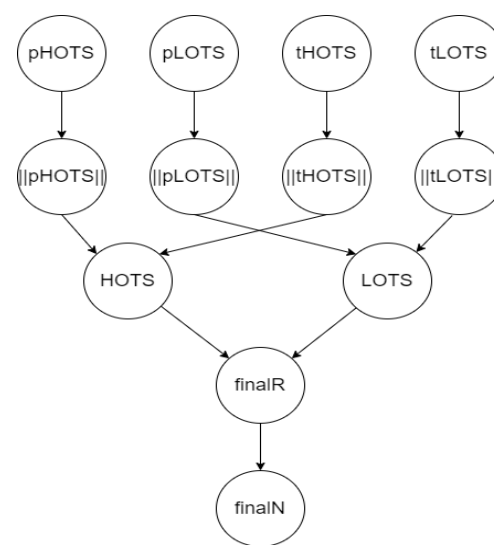
Оценката pHOTS се формира от практическа задача и е в интервала [2, 6]. Точките на останалите оценъчни компоненти се формират при решаване на тест в хартиен или електронен вариант. Крайната оценка finalN е цяло число в скалата за оценяване [2, 6], т.е. $finalN \in [2, 6] \subset N$. Базирайки се на **Модел 1 на йерархична дървовидна организация на компонентите за оценяване**, дефинираме два основни възможни варианта за формализиране логиката на оценителя за оценяване по дисциплината (фиг. 23). При необходимост, чрез дефиниране на допълнителни връзки/зависимости между компоненти на различни нива, може да преминем към **Модел 2**. В практиката при различни ситуации могат да бъдат създадени и експериментирани и други варианти. Важното при всички случаи е, да се определи йерархията от компоненти и връзки между тях, както и множеството от съответните им агрегиращи функции, за даденото оценяване.

Основни компоненти, формиращи крайната оценка при **Вариант 1 за теоретико-практическото йерархично многокомпонентно оценяване** са **теория** и **практика** (фиг. 23 а). Под-компоненти на теорията са **нормализираните форми** на първичните компоненти **tLOTS** и **tHOTS**, а на практиката – нормализираните форми на **pLOTS** и **pHOTS**. Целта на нормализираните форми е свеждането на първичните компоненти до еднаква скала за оценяване (в случая от 2 до 6), така че да бъдат представени до по-близки за оценителя мисловни рамки за оценяването.

От друга страна, при **Вариант 2 за HOTS-LOTS йерархично многокомпонентно оценяване** (фиг. 23 б), за основни компоненти са приети HOTS и LOTS, а техни под-компоненти са съответно нормализираните форми на практическите и теоретични HOTS, и практическите и теоретични LOTS.



Фигура 23.а) **Вариант 1**: Теоретико-практическо йерархично многокомпонентно оценяване за **Модел 1**



Фигура 23.б) **Вариант 2**: HOTS-LOTS йерархично многокомпонентно оценяване за **Модел 1**

И в двата моделирани варианта на първо функционално ниво са нормализирани форми на основните оценъчни компоненти. Конкретните нормализиращи функции могат да следват **либерален** или **строг подход** за оценяване, или тяхна модификация, в зависимост от оценителя. При **либералния подход**, точките от първичните компоненти се свеждат до интервала [2, 6]:

$$(22.1) \quad ||tLOTS|| = 2 + 4 \frac{tLOTS}{30}$$

$$(22.2) \quad ||tHOTS|| = 2 + 4 \frac{tHOTS}{15}$$

$$(22.3) \quad \|pLOTS\| = 2 + 4 \frac{pLOTS}{15}$$

$$(22.4) \quad \|pHOTS\| = pHOTS.$$

При *строгий подход за оценяване*, точките първоначално се изобразяват в интервала [0, 6]. Оценки под 2,5 се закръглят на 2. Съответно, функциите на принадлежност (membership functions) за компонентите от второ ниво са:

$$(23.1.1) \quad \|tLOTS\| = \begin{cases} 2, & 6 \frac{tLOTS}{30} < 2.5 \\ 6 \frac{tLOTS}{30}, & 6 \frac{tLOTS}{30} \geq 2.5 \end{cases}$$

$$(23.2.1) \quad \|tHOTS\| = \begin{cases} 2, & 6 \frac{tHOTS}{15} < 2.5 \\ 6 \frac{tHOTS}{15}, & 6 \frac{tHOTS}{15} \geq 2.5 \end{cases}$$

$$(23.3.1) \quad \|pLOTS\| = \begin{cases} 2, & 6 \frac{pLOTS}{15} < 2.5 \\ 6 \frac{pLOTS}{15}, & 6 \frac{pLOTS}{15} \geq 2.5 \end{cases}$$

$$(23.4) \quad \|pHOTS\| = pHOTS.$$

Представени с IF-THEN логика, формулите изглеждат по следния начин:

$$(23.1.2) \quad \|tLOTS\|: \text{IF } 6 \frac{tLOTS}{30} < 2.5 \text{ THEN } \|tLOTS\| \text{ IS } 2 \\ \text{ELSE } \|tLOTS\| \text{ IS } 6 \frac{tLOTS}{30}$$

$$(23.2.2) \quad \|tHOTS\|: \text{IF } 6 \frac{tHOTS}{15} < 2.5 \text{ THEN } \|tHOTS\| \text{ IS } 2 \\ \text{ELSE } \|tHOTS\| \text{ IS } 6 \frac{tHOTS}{15}$$

$$(23.3.2) \quad \|pLOTS\|: \text{IF } 6 \frac{pLOTS}{15} < 2.5 \text{ THEN } \|pLOTS\| \text{ IS } 2 \\ \text{ELSE } \|pLOTS\| \text{ IS } 6 \frac{pLOTS}{15}$$

Средно-аритметичните стойности от под-компонентите не винаги са удачно решение и затова се стига до субективна преценка от оценителя. Има множество специфични случаи, които е необходимо да се формализират, за да се опише логиката на оценителя. Някои от използваните при нашето оценяване разсъждения за поставяне оценка на компонента *практика* са следните:

- Ако pHOTS или pLOTS е 2, оценката на компонент практика е най-много 3, дори и другия под-компонент да има оценка 6.
- Ако оценката pHOTS е висока, а pLOTS – ниска, то възникват съмнения за нечестни прояви на студента при решаването на задачата, формираща оценката pHOTS. Тогава, поставяме оценката да бъде по-ниска от средно-аритметичната стойност. Това се случва най-често, когато pHOTS е близо до 6, а pLOTS – до 3.
- Ако pLOTS е висока, а pHOTS – ниска, то показаните на теста знания гарантират, че студента има необходимата основа за бъдещо развитие по дисциплината. Въпреки, че не се е представил добре при решаването на задачата, то заслужава стимул в разумни граници. Те трябва да са предварително формализирани с точни числови стойности и правила.

Подобни, разсъждения са валидни и за теоретичните знания. При това, високите оценки tHOTS е добре да се толерират дори и при частичната липса на теоретични знания tLOTS.

Използваните в експеримента правила за поставяне на оценка на практика и теория при *Вариант 1 за теоретико-практическото йерархично многокомпонентно оценяване* са:

(24.1) $\mathit{practice}(\|pLOTS\|, \|pHOTS\|)$:

$$\begin{aligned} & \text{IF } \frac{\|pHOTS\| + \|pLOTS\|}{2} < 2.5 \text{ THEN } \mathit{practice} \text{ IS } 2 \\ & \text{ELSE IF } \|pLOTS\| = 2 \text{ OR } \|pHOTS\| = 2 \text{ THEN } \mathit{practice} \text{ IS } 3 \\ & \text{ELSE IF } (\|pHOTS\| - \|pLOTS\|) \\ & \qquad \geq 2 \text{ THEN } \mathit{practice} \text{ IS } \left(\frac{\|pHOTS\| + \|pLOTS\|}{2} - 0.5 \right) \\ & \text{ELSE } \mathit{practice} \text{ IS } \left(\frac{\|pHOTS\| + \|pLOTS\|}{2} \right) \end{aligned}$$

(24.2) $\mathit{theory}(\|tLOTS\|, \|tHOTS\|)$:

$$\begin{aligned} & \text{IF } \frac{\|tHOTS\| + \|tLOTS\|}{2} < 2.5 \text{ THEN } \mathit{theory} \text{ IS } 2 \\ & \text{ELSE IF } (\|tHOTS\| - \|tLOTS\|) \\ & \qquad \geq 2 \text{ THEN } \mathit{theory} \text{ IS } \left(\frac{\|tHOTS\| + \|tLOTS\|}{2} + 0.3 \right) \\ & \text{ELSE } \mathit{theory} \text{ IS } \left(\frac{\|tHOTS\| + \|tLOTS\|}{2} \right) \end{aligned}$$

За определяне на **крайната оценка** при представения **Вариант 1** е приложено теорията и практиката да са равнопоставени и правилата са формулирани като:

(25) $\mathit{finalR}(\mathit{theory}, \mathit{practice})$:

$$\begin{aligned} & \text{IF } \frac{\mathit{theory} + \mathit{practice}}{2} < 2.5 \text{ THEN } \mathit{finalR} \text{ IS } 2 \\ & \text{ELSE IF } \mathit{theory} = 2 \text{ OR } \mathit{practice} = 2 \text{ THEN } \mathit{finalR} \text{ IS } 3 \\ & \text{ELSE } \mathit{finalR} \text{ IS } \left(\frac{\mathit{theory} + \mathit{practice}}{2} \right) \end{aligned}$$

Крайната оценка finalN се определя като се закръгли finalR към най-близкото цяло число:

(26) $\mathit{finalN}(\mathit{finalR}) = \text{ROUND}(\mathit{finalR})$

Определянето на оценките HOTS и LOTS при **Вариант 2 за HOTS-LOTS йерархично многокомпонентно оценяване** също са обект на подобни разсъждения и могат да бъдат персонализирани в зависимост от вижданията на конкретния оценител.

Формулите за оценка на компонентите **практика** и **теория** (24.1) и (24.2) са съобразени с **Модел 1**, затова участват компоненти единствено от предходното ниво. В определени гранични случаи е удачно използването на подкомпоненти от по-високо ниво и използване на **Модел 2**.

Предимство на размитата логика при оценяване е възможността за постигане на точно автоматизирано оценяване според субективната логика на оценителя. Основен недостатък е трудността за неспециалисти да формализират правилата за формиране на оценката.

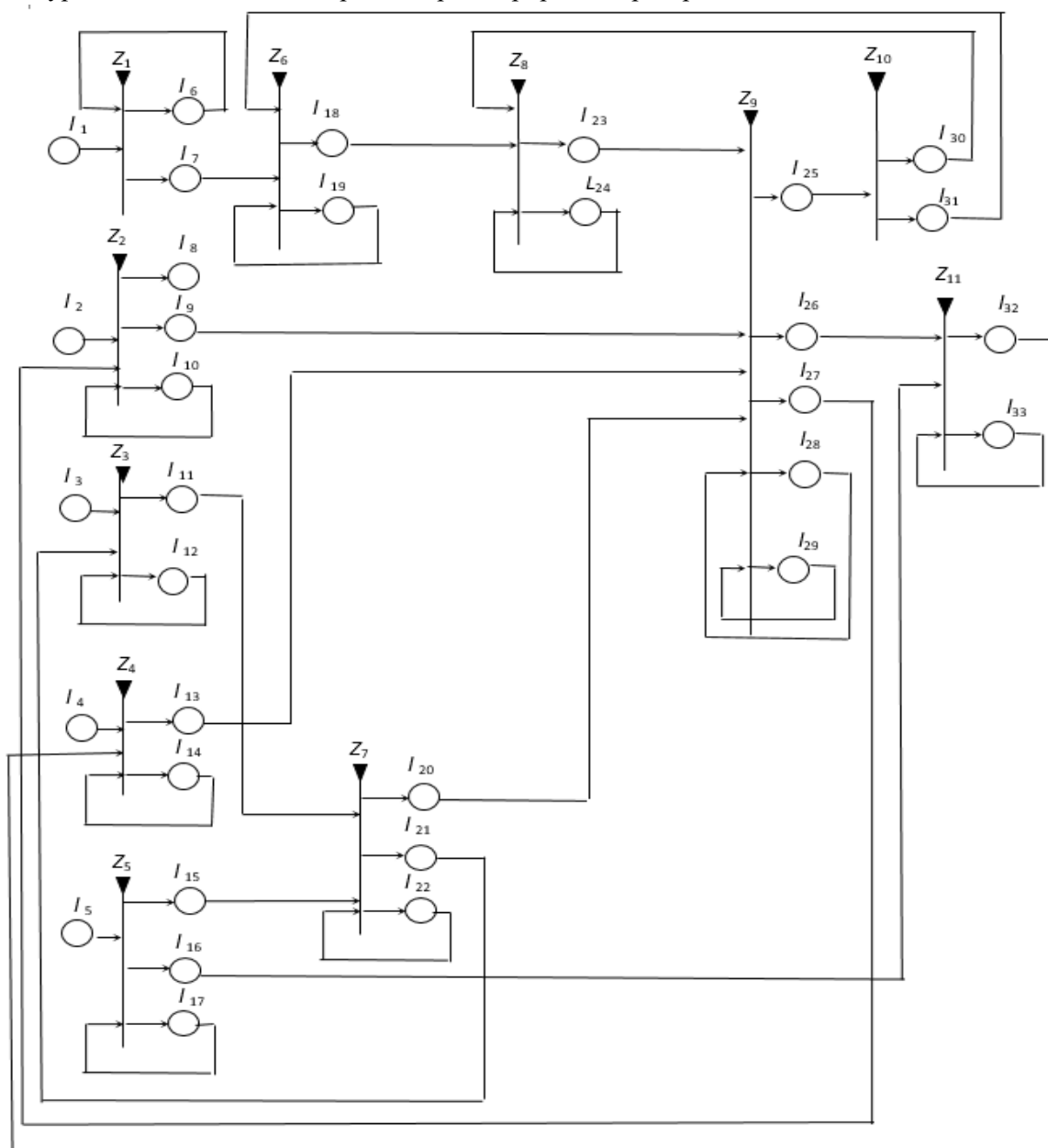
ИЗВОД: използването на размита логика гарантира получаването на коректни оценки, при коректно разписани правила, докато използването на методи на изкуствения интелект изисква достатъчно голям брой подходящи входно-изходни образци за обучение. Следва да отбележим, че всички автоматизирани подходи изискват наблюдение и контрол от страна на преподавателя на формираните оценки. Предложените модели за йерархично многокомпонентно оценяване имат за цел комплексно оценяване на различни мисловни умения от висок и нисък порядък, теоретични знания и практически умения и др. Те позволяват използването на размита логика и могат да бъдат адаптирани за употреба при оценяване на различни дисциплини, в различни възрастови групи.

ГЛАВА 4. ОБОБЩЕНОМРЕЖОВИ МОДЕЛИ ВЪВ ВИРТУАЛНО ОБРАЗОВАТЕЛНО ПРОСТРАНСТВО

4.1. Модел на процесите на персонализиране и използване на среда за обучение

Целта на виртуалното образователно пространство е да интегрира по един „интелигентен“ начин реалния обучаващ процес с виртуално изградения свят. Образователните пространства трябва да предоставят възможности за: активно и интерактивно участие; работа в екип; търсене и споделяне на информация; дискутиране и представяне; произвеждане на нови знания; подпомагане дейностите на обучаемите и обучаващите; връзка с експерти и не на последно място персонализиране на обучението. Инфраструктурата на образователното пространство включва:

- *Изграждащи елементи* на пространството - могат да бъдат: обучаеми, обучаващи, администратори, персонални асистенти, дигитални библиотеки, електронни услуги и т.н.
- *Взаимовръзки* – съществуващите между изграждащите елементи взаимоотношения, осигуряващи съвместната им работа при опериране в пространството.



Фигура 24. OM модел на процесите на персонализиране и използване на среда за обучение

Създадена е серия от модели, свързани с наблюдение и анализ на дейностите на обучаемите в образователна среда, представени в публикациите [60, 70, 72, 97]. Обобщеномрежови модели за работа с големи данни във виртуално колаборативно пространство са създадени в [62, 63, 71].

Тук се представя обобщен модел на процесите на персонализиране и използване на среда за обучение (фигура 24). В този модел студентите (обучаваните обекти) се интерпретират чрез σ -ядра, обучаващите се интерпретират чрез τ -ядра, обслужващият административен персонал (инспектори от учебен отдел) чрез α -ядра, обучителните курсове (SCORM-пакети) в електронната библиотека чрез λ -ядра, електронните услуги предоставяни от средата за електронно обучение чрез Υ -ядра и софтуерните агенти, в ролята на персонални асистенти на обучаемите чрез β -ядра.

Този модел е минимален редуциран обобщеномрежови модел [8, 9]. В дисертационния труд е дадено подробно описание на отделните преходи.

Обобщеномрежовият модел се състои от 11 прехода:

- Z_1 – постъпване на нов студент (обучаем) в електронната среда за обучение;
- Z_2 – постъпване на обучаващ преподавател в електронната среда;
- Z_3 – обслужване на обучението от административен учебен отдел;
- Z_4 – въвеждане на нови курсове за обучение в електронната библиотека;
- Z_5 – работа с електронни услуги, предлагани от обучаващата среда;
- Z_6 – избор на курс за обучение от студент в e-Learning средата;
- Z_7 – процес на администриране на нов курс за обучение;
- Z_8 – работен процес на агента (персонален асистент на студента);
- Z_9 – процес на обучение на студента по избрания електронен курс;
- Z_{10} – процес на оценяване на обучението на студента по избрания курс;
- Z_{11} – процес на изграждане на учебно съдържание на електронен курс.

ИЗВОД: Създаденият ОМ модел [68], дава възможност за проследяване на процесите на персонализация и използване на различни интелигентни инструменти за електронно обучение. Може да се извлече информация за предпочитанията на обучаемите и резултатите от тяхното обучение. На базата на модела и натрупана статистика от реални данни могат да се правят оценки и да се откриват тенденции за развитието на процесите, свързани с електронното обучение и неговото обслужване. Могат да бъдат въведени допълнителни параметри на модела и допълнителни характеристики на ядрата, като се вземат предвид факторите, влияещи върху процеса на обучение, с цел повишаване на интереса на обучаемите. Активното участие на обучаемите в процеса на усвояване на знания и придобиване на умения в голяма степен може да бъде повлияно от качеството на използваното образователно пространство.

4.2. Модел на процеса на прилагане на средства за извличане на знания от данните в среди за обучение чрез апарата на обобщените мрежи

Създаденият обобщеномрежов модел описва възможности за избор и прилагане на подходящи техники за извличане на знания от данните в средите за обучение [59].

Този модел е редуциран обобщеномрежови модел, но от по-висок клас спрямо модела, представен на фиг.24. Тук за всеки преход е дефинирано условие за активиране, като се задава чрез булев израз. Ако стойността му е “*true*”, съответният преход може да се активира, ако е “*false*” – не. Моделът, представен на фиг. 25, съдържа 5 прехода и 21 позиции, групирани в две групи и свързани с два типа ядра, които постъпват в съответните типове позиции: α -ядра и l -позиции представят процеса на извличане на знания от данните, β -ядра и t -позиции представят

критериите за ограничаване на средствата и избора на подходящи техники за извличане на знания от данните. За краткост се използва означението α - и β -ядра вместо α_i - и β_j -ядра, където i, j са номерата на съответните ядра и само влизащите в обобщената мрежа ядра са номерирани с α_0, α_1 и β_1, β_2 . Първоначално едно β_0 -ядро стои в позиция t_6 с начална характеристика:

„налични средства за извличане на знания от данните“.

На следващия преход от функционирането на мрежата, β -ядрото се разделя на две ядра. Оригиналното β_0 -ядро ще продължи да стои в позиция t_6 , докато другото β -ядро ще се придвижи към прехода Z_5 , преминавайки през прехода Z_3 .

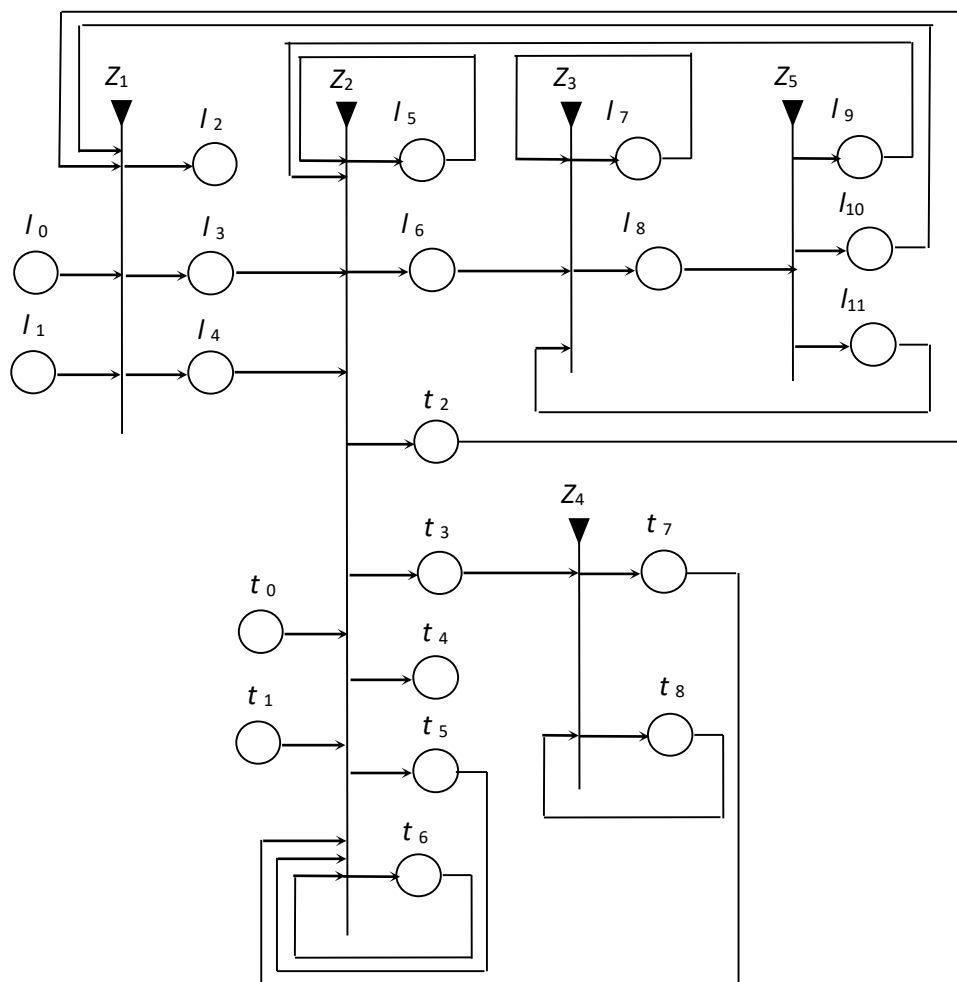
Ядрата α_0 и α_1 , постъпващи в мрежата през позиции l_0 и l_1 имат характеристики съответно:

„начални хипотези“;

„начални данни“.

Ядрата β_1 и β_2 постъпват в мрежата през позиции t_0 и t_1 . Тези ядра получават съответно начални характеристики: „нова техника за извличане на знания от данните“;

„критерии за избор на техника за извличане на знания от данните“.



Фигура 25. OM модел при прилагане на средства за извличане на знания в среда за обучение

ИЗВОД: Интегрирането на системите за обучение със средства за извличане на знания от данните е необходимо за процеса на персонализиране на курсове за електронно и дистанционно обучение. На базата на получените резултати могат да се въведат допълнителни мерки за анализ и промяна на обучаващите курсове и критериите за анализ. Това от своя страна е път към повишаване на качеството на обучението във висшето училище.

4.3. Обобщеномрежов модел на процес за многокомпонентно оценяване

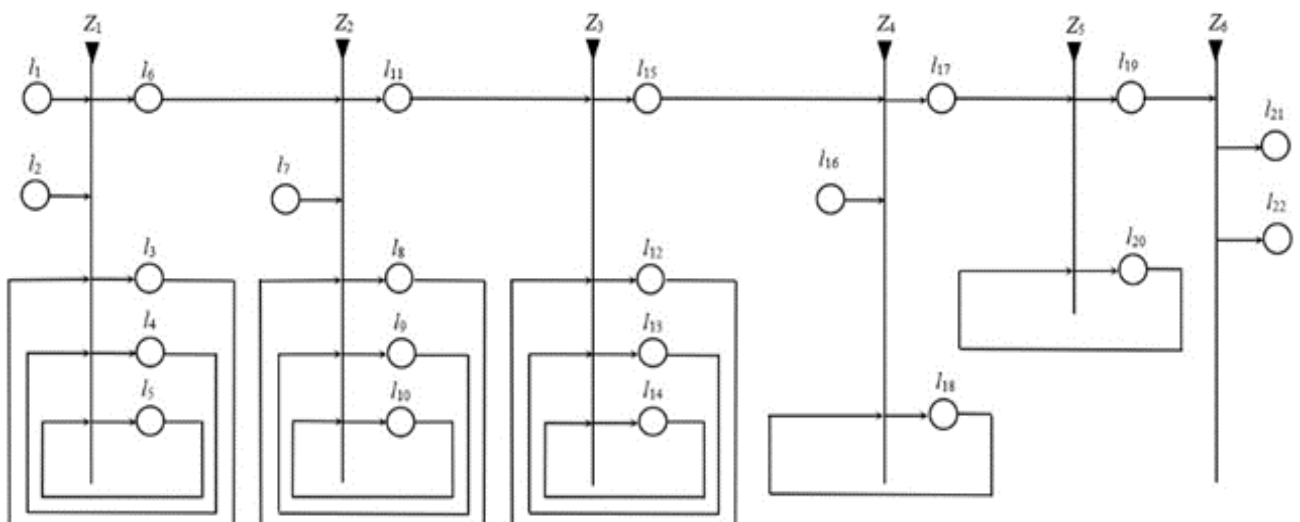
Обективното оценяване на знанията и уменията на обучаемите има множество характеристики: истински резултат (*true score*), надеждност (*reliability*) и валидност (*validity*), честност, диференцираност, всестранност на оценката и др. [91]. То предоставя на преподавателите по-добра възможност както за оценка на усвояването на знанията от обучаемите, така и за анализ на ефективността на преподаването. Създаден е обобщеномрежов **модел на процес за многокомпонентно оценяване**, който включва шест етапа:

1. **Създаване на мета-модел на тест.** Мета-моделът дефинира типове оценъчни компоненти, определящи какви знания и умения се оценяват според конкретна таксономия и моделите за оценяване.
2. **Създаване на модел на тест.** Моделът на тест включва характеристики за целевата група, предмета на обучение, началото и продължителността на теста, и др. На всеки тип оценъчни компоненти на мета-модела се задават съответен брой тестови единици.
3. **Конфигуриране на тест.** Задават се конкретни тестови единици, които се въвеждат или избират от база от данни с въпроси. За всяка конкретна тестова единица се указва към кой тип оценъчни елементи принадлежи. Задават и конкретните обучаеми.
4. **Провеждане на теста** е процес, при който се генерират конкретни тестове. Те включват определения в модела брой тестови единици от всеки тип. Обучаемите въвеждат своите отговори и решения.
5. **Процес на оценяване.** Включва автоматично оценяване на теста според избрания модел и ръчно оценяване на отворените въпроси и задачи, ако има такива.
6. **Анализ на резултатите и оценяване на теста.** Създателят на теста използва получените резултати, за да оцени теста и ефективността на подходите за оценяване.

ОМ моделът, описващ процеса на моделиране и конструиране на тестове за многокомпонентно оценяване на знанията и уменията на обучаемите [43] е представен на фиг.26. Този модел е минимален редуциран обобщеномрежов модел и съдържа множеството от преходи:

$$A = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\}, \text{ където:}$$

- Z_1 – създаване на мета-модел на тест;
- Z_2 – създаване на модел на тест;
- Z_3 – конфигуриране на тест;
- Z_4 – провеждане на тестване;
- Z_5 – процес на оценяване;
- Z_6 – анализ на резултатите и оценяване на теста.



Фигура 26. Обобщеномрежови модел на процес за многокомпонентно оценяване

За описание на процесите в модела са използвани следните ядра:

- α - ядра интерпретират потребителите – администратори, създатели на тестове, автори на въпроси, обучаеми;
- β - ядра – типове оценъчни компоненти, за оценяване на практически или теоретични умения, умения за мислене от нисък или висок порядък, умения, категоризирани според конкретна таксономия за обучение или др.;
- γ - ядра интерпретират модели за оценяване – среднопретеглена оценка на оценъчните компоненти, оценка, базирана на теглови коефициенти, размита логика или др.;
- δ - ядро – мета-модел на тест;
- η - ядро – модел на тест;
- λ - ядро – банка от данни с тестови въпроси;
- μ - ядро – оценъчен компонент (например: тестов въпрос, задача, казус или др.).

Процесът се стартира от администратор на системата, представен чрез α -ядро, който постъпва в мрежата през позиция l_1 с начална характеристика:

„уникален идентификатор и име на потребител на софтуерната система“.

В позиция l_2 се стартира заявка за създаване на нов мета-модел на тест, чрез δ -ядро с начална характеристика:

„уникален идентификатор и наименование на мета-модел на тест“.

В позиция l_3 цикли α -ядро с характеристика:

„списък с налични потребители на софтуерната система“.

В позиция l_4 и в позиция l_5 циклят съответно β -ядро и γ -ядро с характеристики:

„типове оценъчни компоненти“ и

„налични модели за оценяване“.

В дисертационния труд е дадено подробно описание на отделните преходи.

ИЗВОД: Предложен е обобщеномрежови модел, който има за цел от една страна да обобщи процеса на оценяване, а от друга – да даде възможност за персонализация на начина за формиране на тестове и оценяване по време на обучение [43]. Това е постигнато чрез дефиниране на мета-модели и модели за оценяване, които задават рамки за създаване на конкретни тестове и подходи за оценяването им. Разработването на софтуерна система, имплементираща описания процес, ще предостави на преподавателите гъвкава платформа за експериментиране с различни стандартни и авторски подходи за конструиране на тестове и оценяване на обучаемите.

4.4. Обобщеномрежови модел на процесите в проектно-базирано обучение

По своята същност проектно-базираното обучение е педагогически модел на междупредметни дейности, насочени към реално съществуваща проблематика [11]. Основни умения, които обучаемите формират са: да се научат да идентифицират етапите в разработването на проект, да планират дейността си, да спазват планираните срокове, да работят съвместно с други членове на екипа, да оценяват дейността на другите членове на екипа, да самооценяват действията си, да участват в дискусиите по темата на проекта, като формират и защитават аргументирано собствени идеи и умения.

Процес на управление на работата по проекти в обучението

Успешното управление зависи от ръководителя на дисциплината: той може да разпределя задачите за изпълнение, да създава график за работата, да контролира изпълнението на подзадачите и да следи цялостната работа. За целта могат да се правят срещи за представяне на текущите резултати и обсъждане на дейностите или да се работи в подходяща интерактивна среда като: Moodle, Acolad или среда на база Wiki технологията.

(1) Основните дейности при дефиниране на проекти

Основните дейности по подготовката и планирането на проектите се формализират като:

- Спецификация на задачата за конкретен проект.
- Дефиниране на подзадачите, времето, за изпълнение и кой ще се заеме с тях.
- Дефиниране и разпределяне на ресурсите.
- Управление на изпълнението на задачите.
- Събиране на различни данни за статистика и за измерване на развитието на проекта.
- Описание на възможните рискове за проекта и план за управлението му.
- Генериране на различни справки, за работата по проекта.

(2) Планиране на екипи (групи)

Основни дейности, свързани по планирането на групите, работещи по проектите са:

- Определяне на броя на участниците в екип по даден проект.
- Определяне на структурата на екипа.
- Определяне на ръководител (от преподавателя или избран от участниците).

(3) Поставяне на задачата пред студентите

При дефинирането на проекта трябва да бъдат представени пред студентите последователно:

- предварителна информация;
- задачата, поставена интересно и мотивиращо;
- постъпково описание на процеса при изпълнение на задачата;
- набор от препоръчителни информационни източници;
- насоки за организиране и съхраняване на информацията;
- начини и критерии за оценяване на изпълнението на задачата.

(4) Дейности на обучаемите, изпълняващи проекти

При изпълнението на проект, обучаемите се налага да: прецизират поставения въпрос или задача (например да го конкретизират или обобщават); събират и анализират данни от различни източници; споделят, генерират и дискутират различни идеи; правят свои аргументирани предположения, хипотези и предвиждания; провеждат и анализират собствени експерименти; създават артефакти (реферати, бази от данни, мултимедия, модели, прототипи); правят доказателства, обобщения и изводи; съобщават и презентират идеите и откритията си пред други хора; поставят за разглеждане нови въпроси и проблеми.

(5) Проверка и оценяване

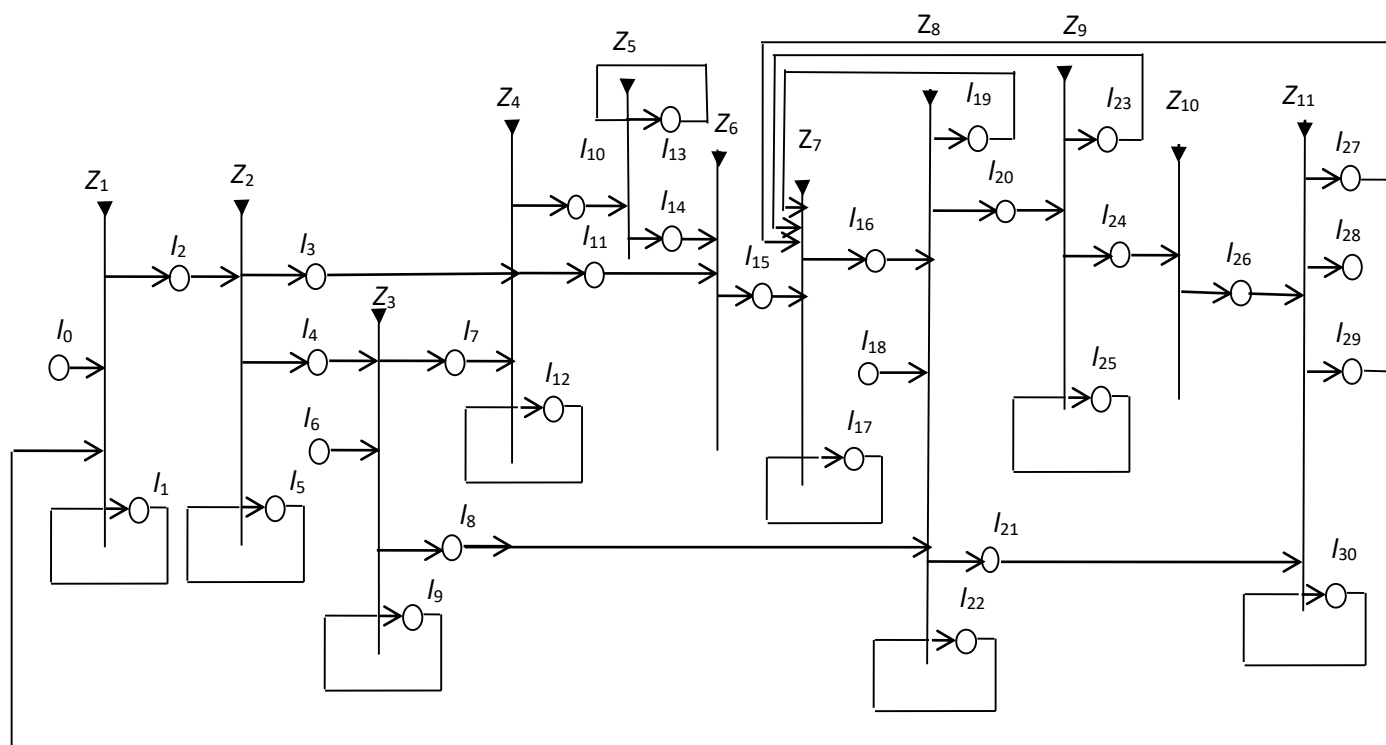
Проверяване и оценяване на резултатите от работата по проекти включва:

- определяне какво точно ще се проверява;
- определяне чрез какви средства ще се извърши проверката;
- коректно цитирани източници;
- публичната изява като естествен завършек за разработвания проект.

Ако преподавателят използва хранилище с данни, в които е събрана информация за минали проекти, то след приключване на работата по текущите проекти, базата от данни трябва да бъде обновена. Натрупваната информация може да се използва за набиране на статистически данни за уменията на студентите и за обучаваната група като цяло.

Моделиране на процеса на обучение чрез работа по проекти

В обобщеномрежовия модел, представен на фиг. 27, задачите (проектите) се интерпретират от α -ядра, преподавателите са интерпретирани от β -ядра, обучаемите (обектите на обучение) от γ -ядра, критериите за оценка съответно чрез δ -ядра [65]. Този модел е минимален редуциран обобщеномрежов модел.



Фигура 27. OM модел на процесите в проектно-базирано обучение

OM моделът се състои от 11 прехода със следното значение:

- Z_1 = Подготовка за проектно-базираното обучение
- Z_2 = Стартира конкретно проектно-базираното обучение
- Z_3 = Формиране на екипи от студенти
- Z_4 = Определяне на задачите за всеки екип
- Z_5 = Дефиниране на подзадачите на всеки проект
- Z_6 = Разпределение на подзадачите по участниците в екипа
- Z_7 = Работа по конкретна подзадача на проекта
- Z_8 = Процес на работа по цялостния проект
- Z_9 = Завършване на работата по проекта
- Z_{10} = Представяне и защита на проекта
- Z_{11} = Оценяване на проектите

В дисертационния труд е дадено подробно описание на отделните преходи.

ИЗВОД: Разгледани са характеристиките и възможностите на проектно-базираното обучение в допълнение на традиционните педагогически подходи. Изгражда се обощеномрежов модел на процеса на обучение, посредством разработване на проект [65]. Моделът служи за анализиране на възможностите за работа по адаптирани към знанията и уменията на студентите проекти, чието разработване води до интегриране на знания, придобивани в различни курсове от обучението. Интеграцията на проектно-базираното обучение с възможностите на електронното и уеб-базираното обучение правят процеса на учене по-привлекателен и по-разнообразен за студентите.

4.5. Модел за игровизация на електронен курс за обучение

Игровизацията на обучението има за цел да интегрира игрови елементи и техники в процеса на е-обучение, за подпомагане усвояването на учебното съдържание. Тези игри имат образователни, тренировъчни или информационни цели. Игровизацията предлага учебният процес да бъде организиран в игрови нива (заклучени или отключени с входни изисквания), които представят различни секции с учебни ресурси и дейности за преминаване. Курсът на обучение се въвежда с интересен контекст (сюжет/история) на дейностите, които трябва да бъдат извършени. Правилата за обучение се въвеждат като игрови правила. Всяко ниво включва предизвикателства - учебни дейности, които трябва да бъдат осъществени за постигане на учебните цели на нивото. Някои от заданията са индивидуални, а други съвместни, където обучаемите работят в екипи.

За завършване на някои дейности обучаемите получават очаквани бонуси (например точки). За постигане на набор от изисквания обучаемите могат да получават значки: (Champion), Суперзвезда (Superstar), Авантюрист (Adventurer), Изследовател (Explorer) и т.н. За отлични резултати или за извършване на конкретна дейност, могат да получат неочаквани награди (допълнителна интересна информация, точки/оценка, материални награди (виртуални предмети) и т.н.). За специфична завършена дейност могат да се получават предимства като комбо (помощ, препоръки, по-подробни примери, удвояване на точки от дадена дейност и т.н.). Някои от учебните елементи могат да са заключени (скрити съкровища) и да бъдат отключени, когато обучаемите отговорят на определени изисквания.

Обучаемите могат да участват в игровизирания курс с конкретна игрова роля, която има и визуално представяне чрез някакъв образ (аватар). Въз основа на събраните точки и текущото игрово ниво обучаемите се подреждат в класацията, където могат да бъдат видени водещите участници в учебния процес. По време на целия процес обучаемите имат достъп до информация за своя напредък в обучението като прогрес в играта и във всеки текущ момент за своя статус (достигнато игрово ниво, спечелени точки, значки и други награди).

Разработени са редица модели за проследяване на процеса на обучение и резултатите на обучаемите [74, 78, 99]. Използването на методи на изкуствения интелект дава възможност за предвиждане на резултатите и частична автоматизация на процеса на оценяване. Обобщеномрежови модел за процесите по избор и изграждане на подходяща E-Test система е даден в [59]. Нов подход за формиране на текущи оценки за знанията и уменията на обучаемите, базирани на размити множества е представен в [39]. Настоящия модел обобщава тези модели и ги разширява с възможности за прилагане на модули за игровизация на обучението и проследяване на обучаемите при работата им с игровизиран курс.

Създаденият обобщен модел на процеса на игровизация на обучаващ курс е минимален редуциран обобщеномрежови модел [69] и съдържа множеството от преходи:

$$A = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\}, \text{ където:}$$

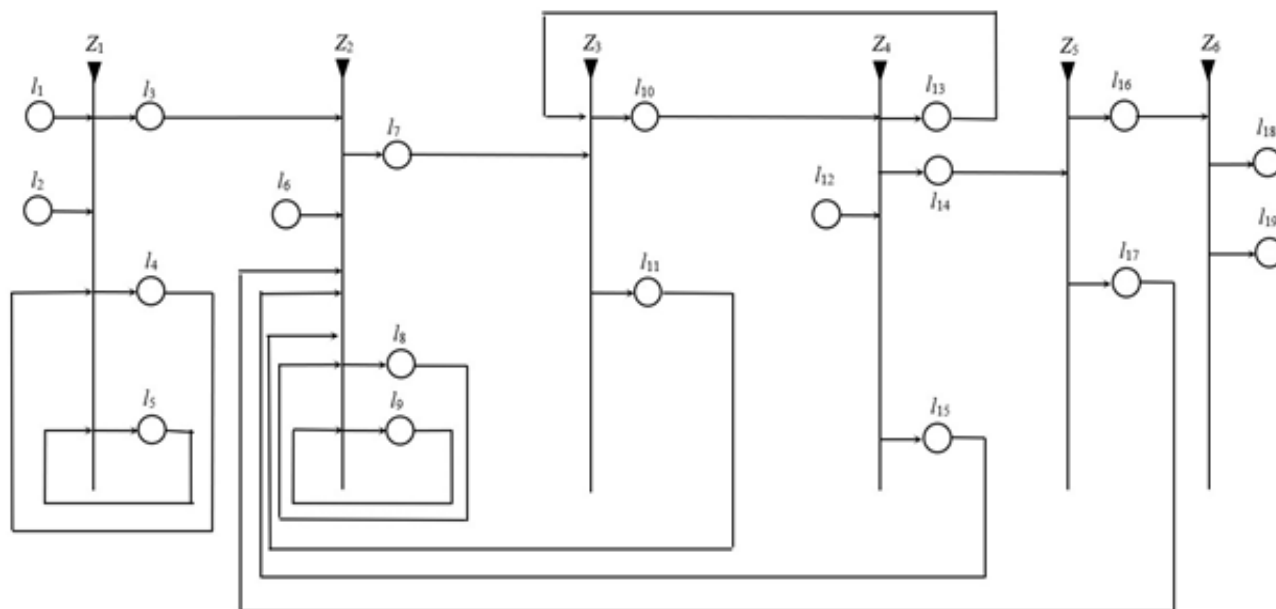
- Z_1 – Създаване на модел и игровизация на обучението;
- Z_2 – Дефиниране на правила за обучението в игровизиран курс;
- Z_3 – Провеждане на обучение с електронния курс;
- Z_4 – Оценяване на обучаемите и награждаване;
- Z_5 – Класиране и анализ на резултатите на обучаемите;
- Z_6 – Анализ на обучението и оценяване на курса.

Следните ядра се използват за описание на процесите:

- α -ядро – потребители на обучаващата среда (администратори, преподаватели, студенти);
- β -ядро – курсове и инструменти за игровизация (плагини, софтуерни компоненти и др.);

- γ -ядро – оценъчни модели;
- η -ядро – данни за обучаваните обекти по време на обучението;
- κ -ядро – специфични данни за игровизираните модули за обучаваните обекти;
- μ -ядро – правила и критерии за прилагане на игрови елементи и награждаване.

Обобщеномрежовият модел на процеса на игровизация на обучаващ курс е представен на фиг. 28. В дисертационния труд е дадено подробно описание на отделните преходи.



Фигура 28. Обобщеномрежови модел на процеса на игровизация на електронен обучаващ курс

Игровизацията на обучението, се разделя на два основни типа: структурна промяна на обучаващ курс и съдържателна промяна на обучаващ курс.

При *структурната промяна* обучаемите преминават през стандартните учебни ресурси, но към курса се включват различни игрови елементи като: точки, значки, класации на първенците, аватари, медали и награди и др.

При *съдържателна промяна* на обучаващ курс се използват игрови техники за представяне на учебното съдържание под формата на правила, нива и умения. Например:

- Правилата на учебния процес могат да бъдат разгледани като правила на игра;
- Мисия/ Предизвикателство/ Приключение – могат да бъдат всички учебни активности, които обучаемият трябва да осъществи в рамките на учебния курс и към тях може да бъде добавен игрови сюжет, който да описва целта на мисията;
- Скрито съкровище - скрити учебни ресурси, които могат да бъдат открити/отворени само при постигане на определени условия (например при завършване на мисия);
- Сюжет / История – учебния процес, може да описва история с мисии за изпълнение.

Игрите притежават силен мотивиращ механизъм за постигане на по-ефективно обучение. Примери за такива среди са: GENIE, The Knowledge Arcade, TalentLMS, Frog, Expertus One, Accord LMS, Axonify, etc. [6, 44]. Тези системи, не са свободно достъпни, а се предлагат като платени софтуери. Средата Мудъл предоставя възможности за създаване на персонализирана учебна среда и доставя набор от инструменти за поддръжка с отворен код, като модули за игровизация могат да се използват: LevelUp, Ranking block и Stash.

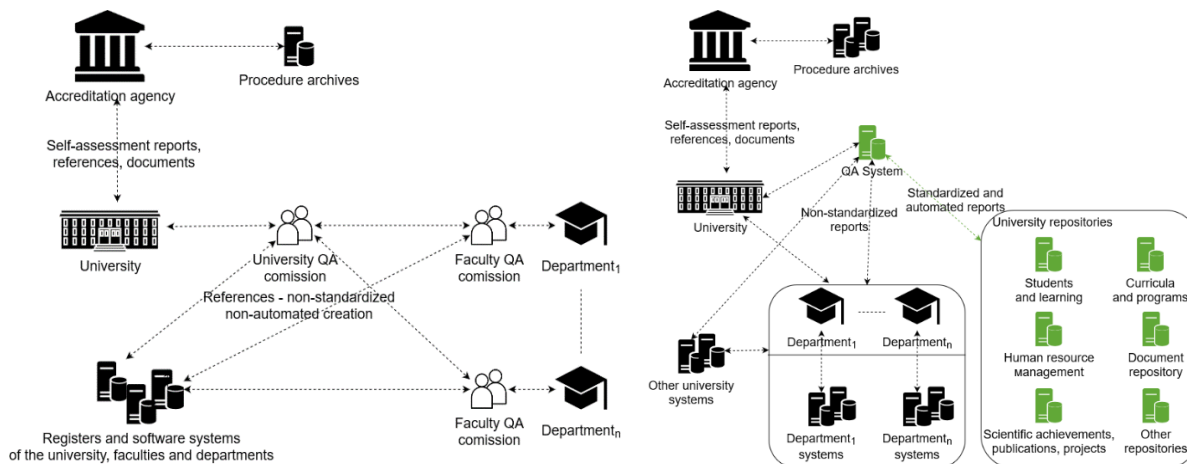
Създаденият обобщеномрежов модел на процеса на игровизация на обучаващ курс предлага решение на един от проблемите, формулиран в книгата на акад. Красимир Атанасов „Generalized Nets and Intuitionistic Fuzziness in Data Mining“ [34].

4.6. Модели за осигуряване на качеството и акредитация на висшето образование

Осигуряването на качеството във висшето образование е непрекъснат процес, който изисква много усилия от страна на образователните институции и акредитиращите организации. Основните форми са две: **вътрешно и външно осигуряване на качеството**. Висшите училища изграждат **вътрешни системи за осигуряване на качеството**, като установяват собствени правила и критерии за оценяване на образователния процес. Те подлежат и на **външно оценяване – акредитация, което се извършва от лицензирани национални и международни агенции**.

Необходимостта от съизмеримост на качеството на висшето образование в Европейския съюз мотивира дефинирането на единни стандарти за качество. През 2005 година са разработени Standards and Guidelines for Quality Assurance in the European Higher Education Area (ENQA). Документа (ESG) [18] съдържа множество стандарти и насоки за прилагането им. България е една от страните, които прилагат стандарта ESG. Осигуряването на качеството на висшето образование в България се извършва от Националната агенция за оценяване и акредитация (НАОА)[2]. **Процедурите за осигуряване на качеството** протичат по обща схема и включват дейностите: самооценяване, посещение на експертна група, оценяване и следакредитационен контрол.

Процедурите за оценяване и акредитация на висшето образование трябва да бъдат естествена и неразделна част от образователния процес и да се провеждат без много усилия, незабележимо за студенти и преподаватели. За целта, е необходимо образователните институции да поддържат хранилища с актуална информация за дейността си, а акредитиращите – да осигурят софтуерни инструменти за автоматизирана оценка на качеството, съгласувани с информацията в националния център за информация и документация (НАЦИД). Модели на текущото състояние са създадени в [40, 43]. Обобщен модел на състоянието е даден на фиг. 29.



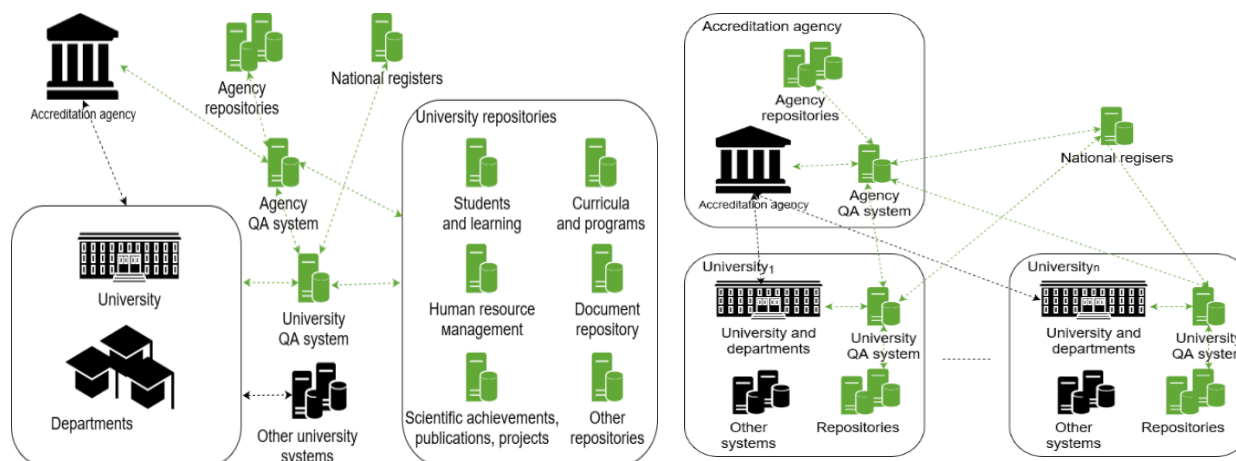
Фигура 29. Модел на процесите при изготвяне на доклад-самооценка от висшите училища

- а) Модел без университетска софтуерна QA система за качество б) Модел с университетска софтуерна QA система за качество

При множество висши училища липсва система, подпомагаща създаването на доклад-самооценка (фиг. 29а). Разпределянето на задачите по изготвянето му става йерархично: НАОА открива процедура за акредитация и осведомява университета; комисията по акредитация разпределя задачи по изготвяне на справки и подготовка на текстове за отделните стандарти; звената събират информация от различни регистри, университетски софтуерни системи, а също и не дигитализирана информация. Създаването на доклада-самооценка е трудно, поради необходимостта от обобщаване на предоставените в различни формат данни и привеждането на текстовете в общ документ. Наличието на инструменти, подпомагащи дейността (фиг. 29б) предоставя възможности за: контрол и проследяване на дейностите; автоматизация на достъпа до справки и др.

4.6.1. Слабо-централизиран модел за акредитация

Един възможен подход за автоматизация на акредитационните процедури е чрез *изграждане на централизирана система на акредитационната агенция и НАЦИД и самостоятелни университетски системи за осигуряване на качеството* (фиг. 30). Всяка университетска QA система автоматизирано извлича необходимите акредитационни данни от университетските хранилища [45]. Друга част от информацията се въвежда от университетски отговорници. В определен момент, информацията от университетската QA система се експортира към QA системата на акредитиращата институция.



Фиг. 30. Слабо-централизиран модел за акредитация

а) Основни системи и комуникационни интерфейси в процеса на акредитация на едно висше училище

б) Основни системи и комуникационни интерфейси в процеса на акредитация в системата на висшето образование

Предимствата на подхода определят и неговите недостатъци. Създаването на собствено решение или поддръжката на стандартизирано приложение от страна на университетите изисква наличието на специализирани ИТ специалисти във всяко висше училище. Създаването на стандартизирана университетска QA система е трудоемка задача, която трябва да се съобрази с наличието на множество различни университетски хранилища и други софтуерни системи.

4.6.2. Силно-централизиран модел за акредитация

Силно-централизираният модел предлага *използването на централизирана cloud система за акредитация*. Системата поддържа модул както за НАОА и НАЦИД, така и конфигурируеми модули, обслужващи акредитационните процеси за отделните висши училища. Използване на единна система предполага улеснена интеграция между основните модули, лесна за потребителите актуализация на софтуера, интеграция на софтуерни разширения, създаване на динамични анализи и справки върху данните, предоставени от всички институции [40].

Основна цел на описаните модели е предлагане на *платформа*, улесняваща работата на всички участници в процедурите по оценяване и акредитация – от страна на обучаващите, и от страна на акредитиращите институции. Използването на софтуерни системи, базирани на предложените модели има много предимства: централизиране на процесите по оценяване, което улеснява тяхното администриране и мониторинг; осигуряване на обща платформа за екипна асинхронна работа на потребители с различни права и отговорности; възможност за работа по всяко време и от всяко място; улеснена комуникация и прозрачност; автоматизирано генериране на документация; оптимизиране на времеви и материални ресурси; намаляване вероятността от допускане на грешки, и др.

4.6.3. Обобщеномрежов модел на процесите на обработка на данни при самооценяване във висше училище

Модел на централизирана система за осигуряване на качеството на висшето образование е създаден в [42]. Обобщеномрежови модел на система за осигуряване на качеството във висше училище е представен в [41]. Тук се създава минимален редуциран обобщеномрежови модел, описващ процесите на обработка на данни при самооценяваща процедура във висшето учебно заведение, който съдържа следния набор от преходи:

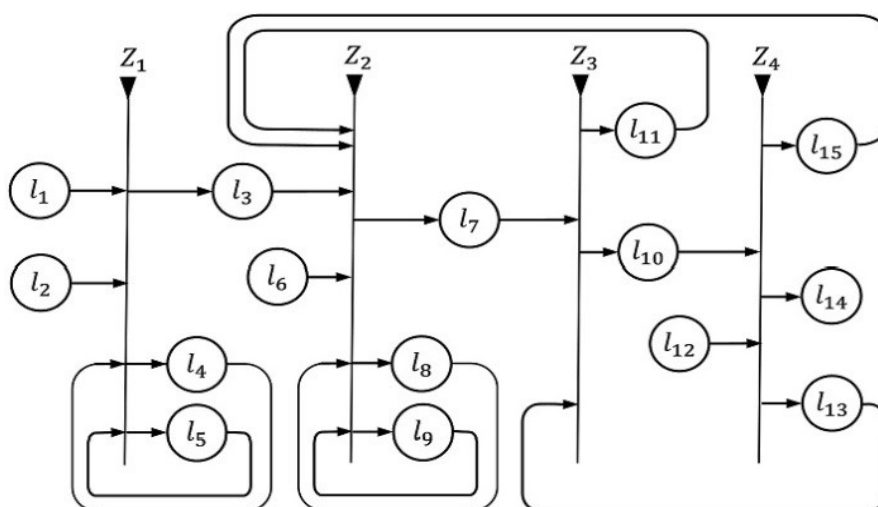
$$A = \{Z_1, Z_2, Z_3, Z_4\},$$

където:

- Z_1 – Събиране на данни от различни източници;
- Z_2 – Интегриране на данни;
- Z_3 – Обработка на данни за изискванията, свързани с критериалната система;
- Z_4 – Изготвяне на доклад за самооценка и съответните приложения.

Следните ядра се използват за описание на процесите:

- α -ядра – данни за дейността на висшето училище;
- β -ядра – членове на административния и академичен персонал, членове на комисията по качеството и групата за изготвяне на доклад за самооценка;
- μ -ядра – критерии и критериална система на акредитиращата институция;
- η -ядро – хранилище на данни с актуална информация за дейността на висшето училище, необходима за изготвяне на доклад за самооценка;
- γ -ядра – инструменти за обработка на данни;
- σ -ядра – заявки за данни от потребител/приложение.



Фигура 31. OM модел на процесите на обработка на данни при самооценяване във ВУЗ

Обобщеномрежовият модел, описващ процесите при обработка на данни при самооценяваща процедура във висшето учебно заведение е представен на фиг. 31. В дисертационния труд е дадено подробно описание на отделните преходи.

ИЗВОД: Предложен е формален модел, описващ процесите на събиране и обработката на данни по процедура за самооценка във висше учебно заведение. Този модел може да бъде развит и чрез използване на **йерархичен оператор** (НЗ от теорията на OM), който замества даден преход или позиция с подмрежа, която има същото, но описано по-подробно поведение на елементите. Въз основа на създадения обобщеномрежов модел и събраните данни от реални процеси могат да бъдат открити поведенчески модели и анализ на производителността на отделните компоненти.

ГЛАВА 5. РЕШЕНИЯ НА НАУЧНОПРИЛОЖНИ ПРОБЛЕМИ В ОБУЧЕНИЕТО

Резултати от приложения на направените изследвания и анализи в образователна среда са представени в няколко авторски публикации. Методически техники за въвеждане на базови концепции в курсове за студенти, адаптирани към знанията и специалностите на обучаемите са предложени в [64, 81]. Как да следваме съвременните тенденции в курсовете за анализ на данни, чрез решаване на подходящо подбрани реални проблеми и задачи се обсъжда в [67]. Подход за представяне на концепции и стандарти за информационна сигурност във висшето образование се предлага в [73]. Програмиране с реактивни блокове се представя в [66] за курсове по Интернет на нещата. В [98] се анализират трудностите при обучението по дедуктивни методи за проверка и синтез на програми. Тенденции в обучението на софтуерни инженери се представят в [34, 35]. В тази глава от дисертационния труд се обобщават резултати от тези публикации и се предлагат основни елементи на учебни програми и подходи за въвеждане на базови модули и инструменти на науката за данните във Висшите училища.

5.1. Учебна програма за анализ на данните

Всяко направление и конкретна специалност се фокусират върху целите, които си поставят с обучението на студентите. Приложните специалности акцентират върху способността на обучаваните да могат да избират подходящ софтуер за анализ на данни. В други специалности акцентът е обучаемите да усвоят основни алгоритми за анализ на данни, машинно обучение и изкуствен интелект. Студентът трябва да може да конструира модели за анализ на съществуваща ситуация и бъдеща прогноза, да се научава как да използва различни техники на изкуствения интелект с цел откриване на аномалии и създаване на оптимални модели. В най-амбициозния план, след завършване на обучението студентът трябва да може да **управлява целия цикъл от данни и да извлича от тях знания с помощта на методи за събиране и анализ на данни и използване на подходящи алгоритми и системи за управление на данните.**

Отчитайки спецификата на обучението в направлението и специалността, учебната програма на дисциплината, свързана с анализ на данните може да акцентира върху различни аспекти. Следва предложение на **базови модули**, които могат да намерят различно място в обучението по дисциплината за анализ на данни (независимо от конкретното наименование).

(1) Предварителна подготовка. Студентите трябва да са преминали основните курсове по програмиране, математика, теория на вероятностите и статистически методи, бази от данни, изкуствен интелект, дискретна математика, структури от данни и комуникация, за да се даде възможност на обучаемия да свързва методологиите с въвежданите нови знания по дисциплината.

(2) Въгъпителна част. В този модул се представят основите етапи в цикъла на анализ и извличане на данни и да се направи преглед на различни инструменти, които поддържат всеки етап от цикъла. Целта на модула е да помогне на студентите в бъдеще успешно да се реализират като анализатори на данни или да провеждат научни изследвания и развиват академична кариера.

(3) Основна част на курса. Този модул трябва да осигури теоретичното обучение с базовите алгоритми и подходи за анализ на данни като: класификационен анализ, регресионен анализ, асоциативен анализ, клъстерния анализ, анализ на шумове и др. Основен акцент в процеса е прилагането на съответен Data Mining алгоритъм или алгоритми, позволяващи получаването на знания, описващи: връзка между свойства на данните, модели на данните, резултатите от класификацията и клъстеризацията на данните и др. Интересно е какви типове зависимости могат да бъдат открити в процеса на извличане на закономерности от данни. За целта се разглеждат шест типа задачи, решаването, на които води до получаване на интересувашите ни зависимости: *описание и обобщение на данни; сегментация (или клъстеризация); анализ на крайностите (outlier detection); описание на понятия (concept description); класификация и регресия; анализ на зависимостите.*

За извличане и представяне на моделите от данни се използват различни математически средства и методи на изкуствения интелект. Основни методи, използвани от техниките за извличане на закономерности от данни са: метод на поддържащите вектори (Support Vector Machines SVM), метод на наивния Бейс (Naïve Bayes), разлагане по метода на главните компоненти (Principal Component Analysis), метод на изкуствена невронна мрежа за анализ на данни (Artificial Neural Networks) и др.

(4) Практическо обучение – обучаемите експериментират с извличане на данни, управление на данни и анализ, подходи за визуализиране на данни, чрез прилагане на различни системи с отворен код. В курса следва да се представят популярни системи за обработване на данни и извличане на знания като: *WEKA* [103], *RapidMiner* [87], *KNIME* [52], *Orange* [56] и др.

(5) Курсов проект. Студентите с по-слаби умения по програмиране и недостатъчни математически знания изпитват трудности при разбирането на същността на анализа на данните и прилагането на избраната техника. За преодоляване на трудностите се предлага като последен модул включване на проектно-базирано обучение. Курсовите проекти - индивидуални или екипни трябва да са със специфична ориентация към определена бизнес област. Например:

- организационния и бизнес анализ на данните: финансови данни, застраховане, банкиране, управление на инвестиции, управление на риска, пазарни данни и др.;

- анализа на данните за общественото здраве: данни за качеството на живота на човека, изграждане на модели за качествен живот, предотвратяване на бъдещи заболявания и епидемии и формулиране на програми за управление на здравето в общността;

- управлението на публичните институции, болници, държавни институции и др.;

- анализ на данни в интелигентни градове и кибер-сигурност.

Независимо дали целта е да се открият интересни взаимовръзки, да се категоризират обекти в групи, да се оптимизира планирането на ресурси или да се определят тарифи за таксуване, работата по проектите ще помогне на студентите да получат основно разбиране на техниките за анализ на данни и да усвоят умения за извличане на знания с конкретни задачи.

ИЗВОД: Развитие на такива академични програми е от решаващо значение за успеха на обучаемите като бъдещи изследователи в дигиталния свят. Интересът на обучаемите и активното им участие в процеса на обучение могат да бъдат повлияни от качеството на обучаващия процес и използваната учебна програма.

5.2. Инструменти за извличане на знания чрез примери

Обучението в курсовете по Базис на данни на студентите от направление *Информатика и компютърни науки* в повечето български университети се провежда през втори или трети семестър. За целта последователно се въвеждат: основни модели на данните; процес на проектиране на релационни бази от данни, анализ и нормализация на релационни схеми, езици за описание и обработка на данни. Студентите се запознават с проблемите при разпределени бази от данни и обектно-ориентирани бази от данни и новите перспективи за развитие на големите данни (BigData). След като реализират редица приложения под формата на курсови проекти, студентите преминават към анализ на хранилища от данни и Data mining техники. Тези техники предлагат богати възможности за изследване на данните, но изучаването им предполага силна математическа основа. Имайки предвид ограниченото време, в курсовете тук се предлага разкриване на възможностите на тези техники, като се предлагат на студентите конкретни примери върху предварително подготвени данни. Студентите могат да анализират данните, следвайки методологията на CRISP-DM модела, като в последователност от шест стъпки се следват шестте основни фази в процеса на извличане на закономерности от данни. За целта се използват подходящо подбрани реални задачи за анализ на данни, които са адаптирани с цел по-лесно прилагане на съответен алгоритъм, подробно представени в дисертационния труд.

ИЗВОД: Анализът на резултатите от проведеното обучение показва, повишаване степента на усвоеност на учебното съдържание и повишаване на интереса към дисциплината [67]. Обучението по тематика не предизвика сериозни трудности сред студентите, които притежават необходимата предварителна подготовка. Студентите с по-слаби умения по програмиране и не достатъчни математически знания изпитват редица трудности при разбирането на анализа и прилагането на избраната техника на извличане на закономерности от данни.

5.3. Въвеждане на онтологии в обучението на студенти

Предлага се метод за въвеждане на онтологии в курсовете на специалностите в направление “Информатика и компютърни науки” чрез реализация на три основни задачи, които се решават в процеса на обучение [64]. Първата задача е свързана с описание на природата на онтологията. Представя се произхода и значението на онтологията и различните видове онтологии. Студентите се запознават с основните компоненти на структурата на онтологията. Втората задача е свързана с представяне на схеми, които реализират онтологии. Тази част от обучението се базира на примери, сравнение между различни примерни онтологии, както и анализ на всяка от тях. За да се направи това, се въвеждат средствата на езика OWL. Третата задача е разработването на пример за OWL онтология, с помощта на софтуера Protégé.

Въвеждането на понятието „онтология“ започва със запознаване на неговия произход от философията и отразява природата на нещата, които реално съществуват. В съвременната литература този термин се използва за **означаване на определена система от категории, които са следствие от определена система от възгледи за света**. Представя се тяхна класификация в зависимост от различни класификационни признаци. В [38] се разглежда класификация, според която онтологията се разделят на: общи онтологии; онтологии, ориентирани към конкретна област; онтологии, ориентирани към конкретна задача и приложни онтологии.

По отношение на структурата на онтологията явно се отделят два компонента: имена на съществуващи концепции и връзки в домейна. Редица ограничения могат да бъдат наложени над домейните. Онтологията заедно с множество конкретни инстанции на класовете съставя база знания (knowledge base). Така разработването на една онтология преминава през следните фази [36]: дефиниране на класовете в онтологията; изграждане на йерархия от класовете; дефиниране на характеристики и описание на възможните стойности за тях; попълване на конкретни стойности за характеристиките, за да се получат конкретни инстанции.

Разработването на онтология е итеративен процес. Основно правило е концепциите в онтологията да са близки до обектите и връзките в конкретния домейн. При построяване на дадена онтология предварително трябва да се определи – с каква цел се създава онтологията, за какви типове въпроси се предполага да бъдат получавани отговори (с нейна помощ), как ще се използва и поддържа. Обучението продължава с изучаване на: методологията за проектиране на онтология; изучаване на Web Ontology Language (OWL) и изграждане на примерна онтология; използване на Reasoner за работа с онтологии. За целите на обучението се разработва пример за OWL онтология със софтуера Protégé.

В резултат се постига въвеждане на онтологията, чрез разясняване на базовите понятия и основните принципи в областта, базирайки се на примери. Разработва се съвместно със студентите примерна онтология, преминавайки през всички стъпки в процеса на създаването. Предоставят се възможности за допълване на изградената онтологичната система – с нови понятия, отношения или атрибути.

ЗАКЛЮЧЕНИЕ

Извличането на данни в образователни среди (Educational Data Mining) и анализът на обучението (Learning Analytics) са сравнително нови области, които имат за цел да подобрят образователния опит, да помагат на заинтересованите страни (преподаватели, студенти, администратори, изследователи), да вземат по-добри решения, използвайки натрупаните данни. Въпреки големите очаквания и нарасналия обем на публикациите в областта на извличането на данни в образователни среди, все още стоят бариери и предизвикателства пред изследователите в областта, като **липсата на изчерпателни и лесни за използване и разбиране инструменти, които могат да бъдат интегрирани в най-популярните среди за управление на обучението.**

Съвременното обучение използва софтуерни платформи за управление на учебното съдържание, които осигуряват технологичен апарат, паралелно с публикуването на учебното съдържание да се събира и съхранява информация и за активността на потребителите – преподаватели и обучаеми. Насока за бъдещо развитие на работата е създаване на лесни за ползване софтуерни инструменти, с възможности за интегриране в средите за управление на обучението за ранното откриване на обучаеми в риск и своевременно уведомяване на преподавателите кои обучаеми се нуждаят от допълнителна помощ и кои практики на преподаване оказват най-голямо въздействие. Освен това трябва да продължат изследванията относно възможностите за персонализиране на учебния процес и съдържание, така че всеки обучаем да получава ресурси според текущите си познания и нагласа за обучаващия процес.

Предизвикателство е разработване на адаптивни курсове, които се персонализират автоматично според профилите на обучаемите (нужди, цели, произход, държава, стил на учене и т.н.). В дисертационния труд фокусът е върху наблюдение и идентифициране на обучаеми в риск, за които има вероятност да отпаднат или да се провалят по време на обучението. Но проведените анализи на данни са върху сравнително малки групи обучаеми. По-добри резултати могат да бъдат получени при анализиране на голям брой студенти, курсове и институции. Например при MOOCs (Massive Open Online Courses) могат да бъдат използвани данни от хиляди студенти. Участниците имат различни произход, зрялост, опит, нива на образование, езикови умения, цели, нужди и стилове на учене. Но значителна бариера, [37] е свързана с **етиката при използване на личните данни, която трябва да се взема предвид на всички етапи - от събирането на данни до тълкуването на резултатите и вземането на решения.** Например данни свързани с пол, социален статус, раса, религиозни вярвания, идеология или увреждане биха могли да доведат до дискриминационно отношение. От друга страна трябва да се вземе предвид и поредицата мерки, свързани със защитата на личните данни.

Следваща стъпка на изследванията в областта е създаване и прилагане на **методологии за използването на Big Data Analytics в образователни среди.** В ерата на големите данни, възможностите за съхраняване, управление и обработка на данни от онлайн среди за обучение дават възможност да се изследват по-добре процесите на обучение и да се търсят ефективни начини за тяхното подобряване. **Комбинацията от Big Data и анализът на обучението (Learning Analytics) е обещаваща област на изследване.**

Приноси на дисертационния труд

А. Научни приноси (в класа обогатяване на съществуващите знания).

Предложени са методи и модели, като резултат от теоретични обобщения на процесите на наблюдение и анализ на дейностите на обучаемите:

1. Петстъпков метод за оценяване и прогнозиране на знанията, уменията и компетентностите на обучаемите във виртуалното образователно пространство, с възможност да се прилагат изчислителни модели и критерии за динамично оценяване, а окончателната оценка се формира от множество оценъчни компоненти. Проведените емпирични изследвания потвърждават, че предложеният метод е перспективен за разработване на система за ранно предупреждение за различни заинтересовани страни от обучаващия процес.
2. Модели чрез размита логика на йерархично многокомпонентно оценяване на различни мисловни умения от висок и нисък порядък, теоретични знания и практически умения: *Модел 1* на йерархична дървовидна организация на оценъчните компоненти и *Модел 2* на йерархична графова (недървовидна) организация на оценъчните компоненти. Представените модели са приложими при различни подходи за оценяване.
3. Метод с уеб метрики и индуктивна размита класификация за оценка на степента на използване на уеб-сайтове от обучаемите, при анализ на поведението им в средата и уеб пространството.
4. Метод за анализ на типа на документ, на базата на класификационни алгоритми, като целевия атрибут е дали документа е свързан с областта на обучение. Анализът на данните дава възможност да се търсят трудностите, които обучаемите срещат при работа с литературни източници и да се проектират модули, които да отговарят на индивидуалните им нужди при търсене и използване на документи.
5. Трифакторен модел на обучаемия, който включва: фактори на компетентност, емоционални фактори, фактори на въздействие на социалната среда. Чрез моделът на обучаемия може да се проследи промяната на знанията, уменията и компетентностите и да се прави предвиждане, според натрупаните данни от работата на среди и системи за електронно обучение.
6. Обобщеномрежови модели: с възможност за проследяване на процесите на използване на различни инструменти в среди за електронно обучение; на процеса на прилагане на Data Mining средства в среди за обучение; на процеса на многокомпонентно оценяване в шест етапа с възможност за персонализация на начина за формиране на тестове чрез дефиниране на мета-модели, задаващи рамки за създаване на тестове и подходи за оценяването им; на процесите в проектното обучение и възможностите за интеграция с електронното и уеб-базираното обучение; за игровизация на E-learning курс, за анализиране на възможностите и проблемите при обучение с игрови ситуации. Създадените модели предлагат частично решение на проблемите, формулиран в книгата на акад. Кр. Атанасов „Generalized Nets and Intuitionistic Fuzziness in Data Mining“ [9].
7. Слабо-централизиран и силно-централизиран модел за осигуряване на качеството и акредитацията във висшето образование. Базов обобщеномрежов модел на информационните потоци при обработката на данни при самооценяване във висшето училище.

Б. Научно-приложни приноси (в класа приложение на научните постижения в практиката)

1. Многостъпкови методи за създаване на модели в образователен контекст чрез инструментите за машинно обучение на системата *Orange Data Mining*.
2. Софтуер на инструменти за анализиране и сравняване на поведението на два агента за машинно обучение: *Rule-Based System* и *Reinforcement Learning*.
3. Софтуерни инструменти за анализ на звукови честоти и преобразуването им в цветове в RGB модела. Модел на система, която реализира преобразуване на звуковите честоти в цвят.
4. Основни модули на учебна програма на науката за данните в дисциплини на Висшите училища.
5. Методически техники за обучение на студенти с инструменти за извличане на знания чрез примери от реални проблеми и задачи за проектиране на онтологии в обучението на студенти.

**Списък на научните публикации по темата на дисертационния труд
в издания, които са реферирани и индексирани в световноизвестни бази данни с научна
информация (Web of Science и Scopus)**

1. **Orozova D.**, Appropriate E-Test System Selection Model, Comptes rendus de l'Academie bulgare des Sciences, 2019, Vol 72, No 6, pp. 811-820, ISSN 1310-1331, DOI: 10.7546/CRABS.2019.06.14, **IF=0.343** (2019), **WoS Q4**, SJR=0.218, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000477784300014>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85070988936&partnerID=MN8TOARS>

2. **Orozova D.**, K. Atanassov, Generalized Net Model of Processes Related to Big Data, Comptes rendus de l'Académie bulgare des Sciences, book No 12, vol. 71, 2018, pp. 1679 -1686 ISSN 13101331, DOI: 10.7546/CRABS.2018.12.13, **IF=0.321** (2018), **WoS Q4**, Scopus SJR=0.205, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000456750000013>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85063227040&partnerID=MN8TOARS>

3. **Orozova D.**, K. Atanassov, Model of Big Data Map/Reduce Processing, Comptes rendus de l'Academie bulgare des Sciences, 2019, Vol 72, No11, pp.1537-1545, ISSN 1310-1331, DOI: 10.7546/CRABS.2019.11.11, **IF=0.343** (2019), **WoS Q4**, SJR=0.218, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000502809100011>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85109014572&partnerID=MN8TOARS>

4. **D. Orozova**, K. Atanassov, M. Todorova, Generalized Net Model of the Process of Personalization and Usage of an e-Learning Environment, Proceedings of the Jangjeon Mathematical Society 19 (2016), No. 4, pp. 615 – 624. ISSN (print):1598-7264, ISSN (online):2508-7916, ISBN:89-87809-15-3, SJR=0.508, (Scopus Q2)

<http://www.scopus.com/inward/record.url?eid=2-s2.0-84996565258&partnerID=MN8TOARS>

5. **Orozova D.**, M. Todorova, Ontology concept in courses on students, TEM Journal. Volume 7, Issue 3, Pages 693-697, ISSN 2217-8309, DOI:10.18421/TEM73-29, 2018. SJR=0.148. (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000442699500029>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85052218825&partnerID=MN8TOARS>

6. Popchev I., **D. Orozova**, Towards Big Data Analytics in the E-learning Space. Cybernetics and Information Technologies, Vol. 19(3), 2019, pp.16-25, ISSN: 1311-9702, DOI: 10.2478/cait-2019-0023, SJR=0.31, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000486349500002>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85073873211&partnerID=MN8TOARS>

7. **Orozova D.**, K. Kaloyanova, M. Todorova, Introducing Information Security Concepts and Standards in Higher Education, TEM Journal. Volume 8, Issue 3, Pages 1017-1024, ISSN 2217-8309, DOI: 10.18421/TEM83-46, 2019, SJR=0.167, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000483978200046>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85072752036&partnerID=MN8TOARS>

8. Popchev I., **D. Orozova**, Towards a multistep method for assessment in e-learning of emerging technologies, Cybernetics and Information Technologies, Volume 20, No 3, Sofia, 2020, pp.116-129, ISSN: 1311-9702, DOI: 10.2478/cait-2020-0032, SJR=0.272, (Scopus Q2)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000572532500009>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85093080750&partnerID=MN8TOARS>

9. Hadzhikolev, E., S. Hadzhikoleva, K. Yotov, **D. Orozova**, Models for Multicomponent Fuzzy Evaluation, with a Focus on the Assessment of Higher-Order Thinking Skills, TEM Journal, Vol.9, No.4, pp.22-28, 2020, ISSN: 2217-8309, DOI: 10.18421/TEM94-43, SJR=0.199, (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000596297300043>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85092629596&partnerID=MN8TOARS>

10. Todorov, J., V. Valkanov, S. Stoyanov, B. Daskalov, I. Popchev, **D. Orozova**. Chapter 6: Personal Assistants in a Virtual Education Space, Sgurev V., Jotsov V., Kacprzyk J. (eds) Practical Issues of Intelligent Innovations. Studies in Systems, Decision and Control, 140, Springer International Publishing AG part of Springer Nature, 2018, Library of Congress Control Number 2011893667, ISSN 2198-4182, DOI: 10.1007/978-3-319-78437-3_6, SJR=0.131, (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000466552900007>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85049662711&partnerID=MN8TOARS>

11. Todorov J., E. Doychev, **D. Orozova**, A. Stoyanova-Doycheva, IoT Multiagent Assistant in Virtual Educational Space, 2018, AIP Conference Proceedings Volume 2048, pp. 020031; ISSN: 0094243X, ISBN: 978-073541774-8, doi: 10.1063/1.5082049, <https://doi.org/10.1063/1.5082049>, SJR=0.182.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000468108800031>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85058786246&partnerID=MN8TOARS>

12. Stoyanova-Doycheva A., T. Glushkova, S. Stoyanov, **D. Orozova**, Lifelong Learning Supported by an Intelligent Tourist Guide, 2018, AIP Conference Proceedings, Volume 2048, pp. 020038; ISSN: 0094243X, ISBN: 978-073541774-8, SJR=0.182, DOI: 10.1063/1.5082056

<https://www.webofscience.com/wos/woscc/full-record/WOS:000468108800038>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85058819893&partnerID=MN8TOARS>

13. Hadzhikolev E., S. Hadzhikoleva, **D. Orozova**, K. Yotov, A Comprehensive Approach to Assessing Higher and Lower Order Thinking Skills, Uncertainty and Imprecision in Decision Making and Decision Support: New Advances, Challenges, and Perspectives, Lecture Notes in Networks and Systems, Volume 338, pp. 164–177, Lecture Notes in Networks and Systems, Volume 338, pp. 164–177, ISSN 2367-3370, 2022, DOI: 10.1007/978-3-030-95929-6_13, SJR=0.151, (Scopus Q4)

<https://www.webofscience.com/wos/woscc/full-record/WOS:000775291100013>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85126177886&partnerID=MN8TOARS>

14. Hadzhikoleva S., **D. Orozova**, N. Andonov, E. Hadzhikolev, Generalized Net Model of a System for Quality Assurance in Higher Education, 45th International Conference Applications of Mathematics in Engineering and Economics (AMEE'19), 2019, pp. 040005-1 -040005-8, ISBN: 978-0-7354-1919-3. AIP Conference Proceedings, 2019, DOI: 10.1063/1.5133515, Part of ISSN: 15517616 0094243X, SJR=0.19.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000521744400033>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85075796767&partnerID=MN8TOARS>

15. Orozov B., **D. Orozova**, An AI application in E-Learning, CEUR Workshop Proceedings 3061, pp. 76-80, September 27–28, 2021, Plovdiv, Bulgaria, ISSN 1613-0073, SJR=0.228 (за 2021).

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85122281151&partnerID=MN8TOARS>

16. Ivan Popchev, **Daniela Orozova**, Algorithms for Machine Learning with Orange System, International Journal of Online and Biomedical Engineering (iJOE), Vol. 19, No. 04, 2023, pp. 109 -123, ISSN: 2626-8493, SJR=0.28, (Scopus Q3), DOI: <https://doi.org/10.3991/ijoe.v19i04.36897>.
<https://www.webofscience.com/wos/woscc/full-record/WOS:000967834800007>
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85152704594&origin=resultslist&sort=plf-f>
17. **D. Orozova**, N. Hristova, An application of analytical data research in e-learning system, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering Volume 1031 (2021) 012058, IOP Publishing, Online ISSN: 1757-899X, DOI: 10.1088/1757-899X/1031/1/012058.
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85101662489&partnerID=MN8TOARS>
18. Popchev I., **D. Orozova**, DataScience: Experience and Trends, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering, Volume 1031 (2021) 012057, IOP Publishing, Online ISSN: 1757-899X, Print ISSN: 1757-8981, doi:10.1088/1757-899X/1031/1/012057.
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85101718405&partnerID=MN8TOARS>
19. M. Todorova, **D. Orozova**, Training Difficulties in Deductive Methods of Verification and Synthesis of Program, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 9, No. 7, 2018, pp. 18-22, ISSN ISSN: 2158107X ISSN (online):2156-5570, doi:10.14569/IJACSA.2018.090703.
<https://www.webofscience.com/wos/woscc/full-record/WOS:000441330600003>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85054031652&partnerID=MN8TOARS>
20. **D. Orozova**, SQL User Queries Execution Model, Advances in Intelligent Systems and Computing book series (AISC, volume 1081), Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives, Volume 1081 AISC, 2021, Pages 184-189, ISSN 2194-5357, DOI: 10.1007/978-3-030-47024-1_20
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85087754542&partnerID=MN8TOARS>
21. Orozov B., **D. Orozova**, Rule Based System Against Reinforcement Learning, CompSysTech'21: ACM International Conference Proceeding Series on Computer Systems and Technologies '21, June 2021, Pages 67–70, ISBN: 978-145038982-2, DOI: 10.1145/3472410.3472437.
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85117607164&partnerID=MN8TOARS>
22. **Orozova D.**, M. Georgieva, A model of the process of Big Data with generalized net, 2016, IEEE 8th International Conference on Intelligent Systems, IS 2016 – Proceedings, 7 2016, Article number 7737487, pp.599-603, ISBN: 978-150901353-1, DOI: 10.1109/SIELA.2016.7543030
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85006035368&origin=resultslist&sort=plf-f>
23. Hadzhikolev E., S. Hadzhikoleva, **D. Orozova**, Digital Model of a Document in a University Document Repository, IEEE XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018, 3-6 June 2018, Bulgaria, ISBN 978-1-5386-3418-9, DOI: 10.1109/SIELA.2018.8447089.
<https://www.webofscience.com/wos/woscc/full-record/WOS:000454837100037>
<http://www.scopus.com/inward/record.url?eid=2-s2.0-85053840788&partnerID=MN8TOARS>

24. **Orozova D.**, A. Ivanov, Generalized Net Model of Virtual Collaboration Space, IEEE XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018, 3-6 June 2018, Bourgas, Bulgaria, ISBN 978-1-5386-3418-9, DOI: 10.1109/SIELA.2018.8447090.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000454837100038>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85053825957&partnerID=MN8TOARS>

25. **Orozova D.**, E. Sotirova, Modeling of a Learning Management System, IEEE XVIII-th International Symposium on Electrical Apparatus and Technologies SIELA 2014, 2014, Bulgaria, pp. 95-98, ISBN: 978-147995817-7, DOI: 10.1109/SIELA.2014.6871879.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000345744800038>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-84906751001&partnerID=MN8TOARS>

26. **Orozova D.**, M. Georgieva, A model of the process of Big Data with Generalized Net, Electrical Apparatus and Technologies (SIELA), 2016, IEEE 19th International Symposium, Bourgas, 238-241, ISSN: 978-1-4673-9521-2, ISBN 978-619-160-648-1, DOI: 10.1109/SIELA.2016.7543030.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000382936800059>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-84985945018&partnerID=MN8TOARS>

27. I. Popchev, **D. Orozova**, S. Stoyanov, IoT and Big Data Analytics in E-Learning, 2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE), IEEE 2019, pp. 1-5, ISBN 978-172816481-6, DOI: 10.1109/BdKCSE48644.2019.9010666.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000563467600027>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85082305981&partnerID=MN8TOARS>

28. **Orozova D.**, I. Popchev, Cyber-Physical-Social Systems for Big Data, IEEE XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, 3-6 June 2020, Bourgas, Bulgaria, pp. 334 – 337, ISBN 978-1-7281-4345-3, DOI: 10.1109/SIELA49118.2020.9167161.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85091338714&partnerID=MN8TOARS>

29. **Orozova D.**, N. Hristova, Generalized Net Model for dynamic decision making and prognoses, IEEE XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, Bulgaria, 2020, pp.330 – 333, ISBN 978-1-7281-4345-3, DOI: 10.1109/SIELA49118.2020.9167077.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85091339727&partnerID=MN8TOARS>

30. Popchev I., **D. Orozova**, Text Mining in the Domain of Plant Genetic Resources, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems, 2020, pp. 596-600, ISBN 978-172815456-5, DOI: 10.1109/IS48319.2020.9200174.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85092747530&partnerID=MN8TOARS>

31. Hadzhikoleva S., **D. Orozova**, E. Hadzhikolev, N. Andonov, Model of a Centralized System for Quality Assurance in Higher Education, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems 2020, pp. 87-92, ISBN 978-172815456-5, DOI: 10.1109/IS48319.2020.9199951.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85092696472&partnerID=MN8TOARS>

32. Georgieva P., E. Nikolova, **D. Orozova**, Data Cleaning Techniques in Detecting Tendencies in Software Engineering, MIPRO 2020, IEEE 43rd International Convention, 2020, Croatia, Opatija, pp.1272-1277, ISSN: 2623-8764, DOI: 10.23919/MIPRO48935.2020.9245416.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000790326400188>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85097229443&partnerID=MN8TOARS>

33. **D. Orozova**, S. Hadzhikoleva, E. Hadzhikolev, A Course Gamification Model for the Development of Higher Order Thinking Skills, MIPRO 2021, IEEE 44th International Convention, September 27, 2021 – October 1, 202, Opatija, Croatia, pp. 1772-1777, ISSN 1847-3946, DOI: 10.23919/MIPRO52101.2021.9597151.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85123052448&partnerID=MN8TOARS>

34. I. Popchev, **D. Orozova**, Data Analytics Approach in Virtual Educational Space, 2021 IEEE Big Data, Knowledge and Control Systems Engineering (BdKCSE), 2021, pp. 1-5, doi: 10.1109/BdKCSE53180.2021.9627247, ISBN 978-166541042-7.

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85123994089&partnerID=MN8TOARS>

35. **D. Orozova**, M. Todorova, Applying Reactive Blocks in Courses of Internet of Things, Proceedings of the 10th Annual International Conference of Education, Research and Innovation, Seville, 16-18 November, 2017, стр. 6960-6967, ISSN (print):2340-1095.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000429975307011>

36. **D. Orozova**, M. Todorova, How to Follow Modern Trends in Courses in "Databases" - Introduction of Data Mining Techniques by Example, Proceedings of the 11th Annual International Technology, Education and Development Conference, Valencia, 2017, 8186-8194, ISSN:2340-1079.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000427401303026>

37. Georgieva P., E. Nikolova, **D. Orozova**, Tendencies in Software Engineering Education, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 3601- 3609, ISBN: 978-84-09-14755-7.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000530109203108>

38. Georgieva P., E. Nikolova, **D. Orozova**, V. Jecheva, Certification In Teaching Information Technologies For Engineering Graduates, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 4564-4573.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000530109204100>

39. D. Tuparova, G. Tuparov, **D. Orozova**, Educational Computer Games and Gamification in the Higher Education – Students’ Points of View, 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, Croatia, Opatija, pp.1879-1884 Electronic ISSN: 2623-8764, DOI: 10.23919/MIPRO48935.2020.9245251.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000790326400284>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85097241244&partnerID=MN8TOARS>

40. Magdalena Todorova, **Daniela Orozova**, Generalized Net Model of Sequential Programs, In proc of the 20th International Symposium on Electrical Apparatus and Technologies (SIELA 2018), 3-6 June 2018, Bourgas, Bulgaria, DOI: 10.1109/SIELA.2018.8447068, ISBN: 978-1-5386-3420-2.

<https://www.webofscience.com/wos/woscc/full-record/WOS:000454837100016>

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85053816562&partnerID=MN8TOARS>

41. Zlatev Z., J. Ilieva, D. Orozova, G. Shivacheva, N. Angelova, Design and Research of a Sound-to-RGB smart acoustic device, Multimodal Technologies and Interaction, 2023, 7(8), 79; WoS, Scopus Q2 SJR=0.504 (за 2022), <https://doi.org/10.3390/mti7080079>

42. **Учебник:** Иван Попчев, **Даниела Орозова**, Представяне на знания в системи с изкуствен интелект, издателство Божич, Бургас, 2018, ISBN 978-619-7181-59-3.

БИБЛИОГРАФИЯ

1. Вълканова, В., М. Сандалски, И. Попчев, С. Стоянов. Образователен портал за подготовка на ученици за зрелостни изпити. - Взаимодействието теория - практика: Ключови проблеми и решения. Международна научна конференция, 2011, 287-294. ISBN: 978-954-9370-80-5.
2. НАОА: Националната агенция за оценяване и акредитация <https://www.neaa.government.bg/>
3. Стоянов С., Д. Орозова, И. Попчев, Е. Дойчев Виртуално пространство за продължаващо обучение, БСУ Научна конференция с международно участие “Хоризонти в развитието на човешките ресурси и знанието” том 2, 2015, стр. 419-425.
4. Годорова М., Подходи, програмни среди и езици за проверка на коректността на програми и прилагането им при подготовката на софтуерни специалисти, София, 2013.
5. Най-добри 12 инструмента и софтуер за уеб анализ за стартиращи фирми <https://squeezegrowth.com/bg/best-web-analytics-tools-for-startups/>
6. Ahlers, R., Garris, R., & Driskell, J. (2014). Games, Motivation, and Learning: A Research and Practice Model. Computer Science Applications. Florida Maxima Corporation.
7. Anderson L. W., D. R. Krathwohl (Eds). A taxonomy for learning, teaching and assessing: A revision of Bloom’s Taxonomy of educational objectives: Complete edition, New York, Longman, 2001.
8. Atanassov K., On Generalized Nets Theory “Prof. Marin Drinov” Publishing House of Bulgarian Academy of Sciences, Sofia, 2007.
9. Atanassov, K., Generalized Nets and Intuitionistic Fuzziness in Data Mining, “Prof. M. Drinov” Publishing House of Bulgarian Academy of Sciences, Sofia, 2020, ISBN 978-619-245-022-9.
10. Barr A., Feigenbaum E., The Handbook of Artificial Intelligence, vol.2, 1984.
11. Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Retrieved from <http://tech.ed.gov/wp-content/uploads/2014/03/edm-ia-brief.pdf>, последен достъп: 01.02.2023
12. Big Data Definition, Teradata, <https://www.teradata.com/Glossary/What-is-Big-Data>: 01.03.2023
13. Bloom B., M. Engelhart, E. Furst, et. al., “Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain”. New York: David McKay Company, 1956.
14. Bradley, J., Atkins, E. Coupled Cyber-Physical System Modeling and Coregulation of a CubeSat, IEEE Transactions on Robotics, 2015, Vol. 31, No. 2, 443-456.
15. Brookhart C., “How to Assess Higher-Order Thinking Skills in Your Classroom”, publ. ASCD, 2010, ISBN: 978-1-4166-1048-9.
16. Buckland, B., F. Schreier & TH. Winkler. Democratic Governance Challenges of Cyber Security. DCAF Horizon 2015 Working No. 1, Geneva Centre for the Democratic Control of Armed Forces.
17. Egger, N. etc., Themes in Science & Technology Education, 2014, Vol. 7 Issue 1, p.3-17, 15p.
18. Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG). (2015). Brussels, Belgium. ISBN: 978-9-08-168672-3.
19. Bower, J., C. Christensen. Disruptive Technologies: Catching the Wave. Harvard Business Review, January-February 1995, pp. 43-53.
20. Brynjolfsson E., A. McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, W.W. Norton & Company, 2014.
21. Cao L. Data science: a comprehensive overview. ACM Comput Surv (CSUR). 2017;50 (3) pp.1-42.
22. Chapman A., “Principles and Methods of Data Cleaning - Primary Species and Species-Occurrence Data.”, Report for the Global Biodiversity Information Facility, Copenhagen, 2005.
23. Chatti, M, A, Dychhoff, A, L, Schroeder, U., and Thus, H. A reference model for learning analytics, International journal of Technology Enhanced learning, 2012.
24. Chemchem, A., Drias, H. From data mining to knowledge mining: Application to intelligent agents. International Joint Conference on Artificial Intelligence: Expert Systems with Applications, (2014).
25. Clifton, B., 2008: Advanced Web Metrics with Google Analytics, Wiley, New York, USA.
26. Dietrich D., B. Heller, B. Yang, Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Published John Wiley & Sons, Inc., 2015, ISBN: 978-1-118-87613-8

27. Ekman P., W.V. Friesen, Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 1971, 17(2), 124–129.
28. Escobar-Jeria, V. H., Martín-Bautista, M. J., Sánchez, D., Vila, M., 2007: Web Usage Mining Via Fuzzy Logic Techniques. In: Melin, P., Castillo, O., Aguilar, I. J., Kacprzyk, J., Pedrycz, W. (Eds.), 2007: *Lecture Notes In Artificial Intelligence*, Vol. 4529, Springer, New York, USA, pp. 243-252.
29. Estivill – Castro, V. & Yang, J. A. (2000). Fast and robust general purpose clustering algorithm. *Pacific Rim International Conference on Artificial Intelligence* (pp. 208-218).
30. Etheridge, D., *Excel Programming: Your Visual Blueprint for Creating Interactive Spreadsheets*, John Wiley & Sons, Inc. 2010.
31. Ethics Guidelines for Trustworthy AI [Online]. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> последен достъп: 01.03.2023
32. Fraley C., Raftery A. E. (1998). How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. Technical Report No. 329. Department of Statistics University of Washington.
33. Georgieva P., E. Nikolova, D. Orozova, Data Cleaning Techniques in Detecting Tendencies in Software Engineering, MIPRO 2020, 43rd International Convention, 2020, Croatia, Opatija, pp.1272-1277, Electronic ISSN: 2623-8764
34. Georgieva P., E. Nikolova, D. Orozova, Tendencies in Software Engineering Education, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 3601- 3609, ISBN: 978-84-09-14755-7
35. Georgieva P., E. Nikolova, D. Orozova, V. Jecheva, Certification In Teaching Information Technologies For Engineering Graduates, ICERI2019, 12th International Conference of Education, Research and Innovation, Seville (Spain) 11-13 November, 2019, pp. 4564-4573
36. Gomez-Perez A., M. Fernando-Lopez, O. Corcho. “Ontological Engineering with examples from the areas of Knowledge management,” *e-Commerce and the Semantic Web 2nd. Ed.* London, Springer-Verlag, 2004.
37. Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics, *Educational Technology & Society*, 15(3), 42-57.
38. Gruber R., Translation Approach to Portable Ontologies, *Knowledge Acquisition*, 1993 V.5. 199-220.
39. Hadzhikolev, E., S. Hadzhikoleva, K. Yotov, D. Orozova, Models for Multicomponent Fuzzy Evaluation, with a Focus on the Assessment of Higher-Order Thinking Skills, *TEM Journal*, Vol.9, No.4, pp.22-28, 2020.
40. Hadzhikolev E., S. Hadzhikoleva, D. Orozova, Digital Model of a Document in a University Document Repository, XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018.
41. Hadzhikoleva S., D. Orozova, N. Andonov, E. Hadzhikolev, Generalized Net Model of a System for Quality Assurance in Higher Education, 45th International Conference Applications of Mathematics in Engineering and Economics (AMEE'19), Sozopol, June, 2019, pp. 040005-1 -040005-8, ISBN: 978-0-7354-1919-3. AIP Conference Proceedings.
42. Hadzhikoleva S., D. Orozova, E. Hadzhikolev, N. Andonov, Conceptual Model of a Centralized System for Quality Assurance in Higher Education, *Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems 2020*, pp. 87-92, ISBN 978-172815456-5, DOI 10.1109/IS48319.2020.9200174.
43. Hadzhikolev E., S. Hadzhikoleva, D. Orozova, K. Yotov, A Comprehensive Approach to Assessing Higher and Lower Order Thinking Skills, Uncertainty and Imprecision in Decision Making and Decision Support: New Advances, Challenges, and Perspectives, *Lecture Notes in Networks and Systems*, Volume 338, pp. 164–177, *Lecture Notes in Networks and Systems*, Vol. 338, pp. 164–177.
44. Hayashi, C. (1998). What is Data Science ? Fundamental Concepts and a Heuristic Example. In: Hayashi, C., Yajima, K., Bock, HH., Ohsumi, N., Tanaka, Y., Baba, Y. (eds) *Data Science*,

- Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Tokyo.
45. Han J., Kamber M., Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2001.
 46. Hornick M., E. Marcade, S. Venkayala, Java data mining: strategy, standard and practice, A practical Guide for Architecture, Design, and Implementation, 2006.
 47. International Educational Data Mining Society. 2022, <https://educationaldatamining.org>. последен достъп: 01.03.2023
 48. Jecheva V., D. Orozova, Ontology-Based Electronic Test Result Evaluation, Advances in Intelligent and Soft Computing, Third International Conference of Software, Services and Semantic Technologies S3T, Springer, 213-214, 2011. ISSN: 1867-5662.
 49. Kapp, K. (2012). The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education. Pfeiffer.
 50. Kaufmann, M., Meier, A., 2009: An Inductive Fuzzy Classification Approach applied to Individual Marketing, In: Proceedings of the 28th North American Fuzzy Information Processing Society Annual Conference, USA.
 51. Kaushik, Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity, Wiley, 2010
 52. KNIME – Open for Innovation: <https://www.knime.com/> последен достъп: 01.03.2023.
 53. Marr B., “Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance”. John Wiley & Sons Ltd, 2015.
 54. Millington, Ian, 2006. Artificial Intelligence for Games. Elsevier Inc. Ch.7.6 Reinforcement Learning, 612-628.
 55. Mobasher, B., 2007: Web Usage Mining, In: (Liu 2007, pp. 449-483).
 56. Orange system: <https://orange.biolab.si/training/introduction-to-data-mining/>, достъп: 01.03.2023
 57. Orozov B., D. Orozova, An AI application in E-Learning, CEUR Workshop Proceedings 3061, pp. 76-80, September 27–28, 2021, Plovdiv, Bulgaria, ISSN 1613-0073.
 58. Orozov B., D. Orozova, Rule Based System Against Reinforcement Learning, CompSysTech'21: International Conference on Computer Systems and Technologies '21, June 2021, Pages 67–70, ISBN: 978-145038982-2
 59. Orozova D., Appropriate E-Test System Selection Model, Comptes rendus de l'Academie bulgare des Sciences, 2019, Vol 72, No 6, pp. 811-820, ISSN 1310-1331.
 60. Orozova D., SQL User Queries Execution Model, Advances in Intelligent Systems and Computing book series (AISC, volume 1081), Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives, pp.184-189, 2020, ISSN 2194-5357, DOI: 10.1007/978-3-030-47024-1_20
 61. Orozova D., I. Popchev, Cyber-Physical-Social Systems for Big Data, XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, pp. 334 – 337, ISBN 978-1-7281-4345-3
 62. Orozova D., K. Atanassov, Generalized Net Model of Processes Related to Big Data, Comptes rendus de l'Académie bulgare des Sciences, book No 12, vol. 71, 2018, pp. 1679 -1686.
 63. Orozova D., K. Atanassov, Model of Big Data Map/Reduce Processing, Comptes rendus de l'Academie bulgare des Sciences, 2019, Vol 72, No11, pp.1537-1545.
 64. Orozova D., M. Todorova, Ontology concept in courses on students, TEM Journal. Volume 7, Issue 3, Pages 693-697, ISSN 2217-8309, DOI: 10.18421/TEM73-29, 2018. ISSN (print):2217-8309.
 65. Orozova D., M. Krawczak, Project-based Learning Modelled with a Generalized Net, Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Warsaw, Poland, Vol.12, 2015/2016, pp. 68–79, ISBN 978-83-61551-13-3.
 66. Orozova D., M. Todorova, Applying Reactive Blocks in Courses of Internet of Things, Proceedings of the 10th Annual International Conference of Education, Research and Innovation, Seville, 16-18 November, 2017, стр.6960-6967, ISSN (print):2340-1095
 67. Orozova D., M. Todorova, How to Follow Modern Trends in Courses in "Databases" - Introduction of

- Data Mining Techniques by Example, Proceedings of the 11th Annual International Technology, Education and Development Conference (INTED), Valencia, 2017, стр.8186-8194, ISSN:2340-1079.
68. Orozova D., K. Atanassov, M. Todorova, Generalized Net Model of the Process of Personalization and Usage of an e-Learning Environment, Proceedings of the Jangjeon Mathematical Society 19 (2016), No. 4, 615 – 624. ISSN (print):1598-7264, ISSN (online):2508-7916, ISBN:89-87809-15-3.
69. Orozova, S. Hadzhikoleva, E. Hadzhikolev, A Course Gamification Model for the Development of Higher Order Thinking Skills, MIPRO 2021, 44th International Convention, September 27, 2021 – October 1, 202, Opatija, Croatia, pp. 1772-1777, ISSN 1847-3946.
70. Orozova D., E. Sotirova, Modeling of a Learning Management System, XVIII-th International Symposium on Electrical Apparatus and Technologies, Proceedings of digests, SIELA 2014, 29-31 May 2014, Bulgaria, pp. 95-98, ISBN: 978-147995817-7
71. Orozova D., M. Georgieva, A model of the process of Big Data with Generalized Net, Electrical Apparatus and Technologies (SIELA), 2016 19th International Symposium, Bourgas, 238-241, ISSN: 978-1-4673-9521-2, ISBN 978-619-160-648-1
72. Orozova D., A. Ivanov, Generalized Net Model of Virtual Collaboration Space, XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018, 3-6 June 2018, Bourgas, Bulgaria, ISBN 978-1-5386-3418-9
73. Orozova D., K. Kaloyanova, M. Todorova, Introducing Information Security Concepts and Standards in Higher Education, TEM Journal. Volume 8, Issue 3, Pages 1017-1024, ISSN 2217-8309, DOI: 10.18421/TEM83-46, August 2019.
74. Orozova D., N. Hristova, Generalized Net Model for dynamic decision making and prognoses, XXI International Symposium on Electrical Apparatus and Technologies SIELA 2020, 3-6 June 2020, Bourgas, Bulgaria, pp. 330 – 333, ISBN ISBN 978-1-7281-4345-3.
75. Orozova D., N. Hristova, An application of analytical data research in e-learning system, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering Volume 1031 (2021) 012058, IOP Publishing, Online ISSN: 1757-899X
76. Petri C.A., “Kommunikation mit Automaten”. Bonn:Institut fuer Instrumentelle Mathematik, Schriften des IIM Nr.3, 1962. Also, English translation: “Communication with Automata”, New York: Griffiss Air Force Base. Tech. Rep. RADC-TR-65-377, vol.1, Suppl. 1, 1966.
77. Popchev I., D. Orozova, Towards a multistep method for assessment in e-learning of emerging technologies, Cybernetics and Information Technologies, Vol. 20, 3, 2020, 116-129, ISSN: 1311-9702
78. Popchev I., D. Orozova. Towards Big Data Analytics in the E-learning Space. Cybernetics and Information Technologies, Vol. 19(3), 2019, pp. 16-25. ISSN: 1311-9702.
79. Popchev, D. Orozova, S. Stoyanov, IoT and Big Data Analytics in E-Learning, 2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE), 2019, pp. 1-5, ISBN 978-172816481-6
80. Popchev I., D. Orozova, Text Mining in the Domain of Plant Genetic Resources, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems, 2020, pp. 596-600.
81. Popchev I., D. Orozova, DataScience: Experience and Trends, ICTTE 2020, International Conference on Technics, Technologies and Education, Journal IOP Conference Series: Materials Science and Engineering, Volume 1031 (2021) 012057, IOP Publishing, Online ISSN: 1757-899X, Print ISSN: 1757-8981 doi:10.1088/1757-899X/1031/1/012057.
82. Popchev I., D. Orozova, Data Analytics Approach in Virtual Educational Space, Big Data, Knowledge and Control Systems Engineering (BdKCSE), 2021, pp. 1-5, ISBN 978-166541042-7.
83. Popchev I., D. Orozova, Algorithms for Machine Learning with Orange System, International Journal of Online and Biomedical Engineering (iJOE), Vol. 19, No. 04, 2023, pp. 109 -123, ISSN: 2626-8493, DOI: <https://doi.org/10.3991/ijoe.v19i04.36897>
84. Popchev, I., I. Radeva, Risk Analysis – an Instrument for Technology Selection. Engineering Sciences (2019) LVI (4) 5-20, ISSN 1312-5702, DOI: 10:7546/ Eng.Sci.LVI.19.04.01.
85. Policy and Investment Recommendations for Trustworthy Artificial Intelligence [Online]

- <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence> последен достъп: 01.06.2023
86. Popper, Karl Raimund (1972). Objective knowledge. Oxford, Clarendon Press
 87. RapidMiner - Depth for Data Scientists: <https://rapidminer.com/>
 88. Reinforcement Learning, Searching for optimal policies: <https://www.cs.upc.edu/~mmartin/Ag4-4x.pdf>
 89. Russell, S and P. Norvig. Artificial Intelligence: A Modern Approach, Prentice Hall, Upper Saddle River, New Jersey, 1995.
 90. Siemens, G, and Long, P, Penetrating the fog-analytics in learning and education. Asynchronous Learning Networks, 2011.
 91. Schaughency E., J. Smith, J. van der Meer, and D. Berg, Classical Test Theory and Higher Education: Five Questions, In: Handbook on Measurement, Assessment, and Evaluation in Higher Education, Edited by Charles Secolsky and D. Brian Denison, Publ. Routledge, 2017, ISBN: 1138892157.
 92. Social Analytics 2022, <https://www.gartner.com/en/information-technology/glossary/social-analytics>
 93. Stoyanov, S., I. Popchev, I., E. Doychev, D. Mitev, V. Valkanov, A. Stoyanova-Doycheva, V. Valkanova, I. Minov. DeLC Educational Portal. – Cybernetics and Information Technologies, Vol. 10, 2010, No. 3, 49-69, ISSN: 1311-9702
 94. Stoyanova-Doycheva A., T. Glushkova, S. Stoyanov, D. Orozova, Lifelong Learning Supported by an Intelligent Tourist Guide, 2018, AIP Conference Proceedings, Volume 2048, pp. 020038; ISSN: 0094243X, ISBN: 978-073541774-8, DOI: 10.1063/1.5082056
 95. Todorov, J., V. Valkanov, S. Stoyanov, B. Daskalov, I. Popchev, D. Orozova. Chapter 6: Personal Assistants in a Virtual Education Space, Sgurev V., Jotsov V., Kacprzyk J. (eds) Practical Issues of Intelligent Innovations. Studies in Systems, Decision and Control, 140, Springer International Publishing AG part of Springer Nature, 2018, ISSN 2198-4182, ISSN 2198-4190 (electronic), DOI:10.1007/978-3-319-78437-3, 22, 131-153
 96. Todorov J., E. Doychev, D. Orozova, A. Stoyanova-Doycheva, IoT Multiagent Assistant in Virtual Educational Space, AIP Conference Proceedings, Vol. 2048, 020031; ISSN: 0094243X,
 97. Todorova M., D. Orozova, Generalized Net Model of Sequential Programs, In proc of the 20th International Symposium on Electrical Apparatus and Technologies (SIELA 2018), 2018, Bulgaria, DOI: 10.1109/SIELA.2018.8447068, ISBN: 978-1-5386-3420-2.
 98. Todorova M., D. Orozova, Training Difficulties in Deductive Methods of Verification and Synthesis of Program, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 9, No. 7, 2018, pp. 18-22, doi:10.14569/IJACSA.2018.090703.
 99. Tuparova D., G. Tuparov, D. Orozova, Educational Computer Games and Gamification in the Higher Education – Students’ Points of View, 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, Croatia, pp.1879-1884 ISSN: 2623-8764.
 100. Unsupervised Learning and Data Clustering: <https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a>, последен достъп: 01.03.2023
 101. Venkatram K., Geetha M. A., Review on Big Data & Analytics– Concepts, Philosophy, Process and Applications, Cybernetics and Information Technologies, Vol. 17(2), 2017, 3-27, ISSN: 1311-9702;
 102. Wang F-Y., Driving into the future with ITS, IEEE Intelligent System, Vol.21, No.3, 2006, 94–95.
 103. WEKA Waikato Environment for Knowledge Analysis: <http://www.cs.waikato.ac.nz/ml/weka>
 104. Wordclouds - free online generator: www.wordclouds.com/
 105. Zadeh, L., The concept of a linguistic variable and its application to approximate reasoning, Published 1975, Information Sciences, v. 8, pp. 199–249.
 106. Zlatev Z., J. Ilieva, D. Orozova, G. Shivacheva, N. Angelova, Design and Research of a Sound-to-RGB smart acoustic device, Multimodal Technol. Interact. 2023, 7(8), 79.
 107. Zumstein D., M. Kaufmann, A Fuzzy Web Analytics Model for Web Mining, IADIS European Conference on Data Mining, 2009, pp.59-66, ISBN: 978-972-8924-88-1.

Abstracts of Dissertations

Number 9, 2023

INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES
BULGARIAN ACADEMY OF SCIENCES

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ИНСТИТУТ ПО ИНФОРМАЦИОННИ И КОМУНИКАЦИОННИ ТЕХНОЛОГИИ

Брой 9, 2023

Автореферати на дисертации