BULGARIAN ACADEMY OF SCIENCES
INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Daniela Ananieva Orozova

Application of data science in the virtual educational space

ABSTRACT

of a dissertation
to acquire the scientific degree "Doctor of Sciences"

by professional direction 4.6. Informatics and Computer Science

Sofia, 2023

CONTENTS

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.

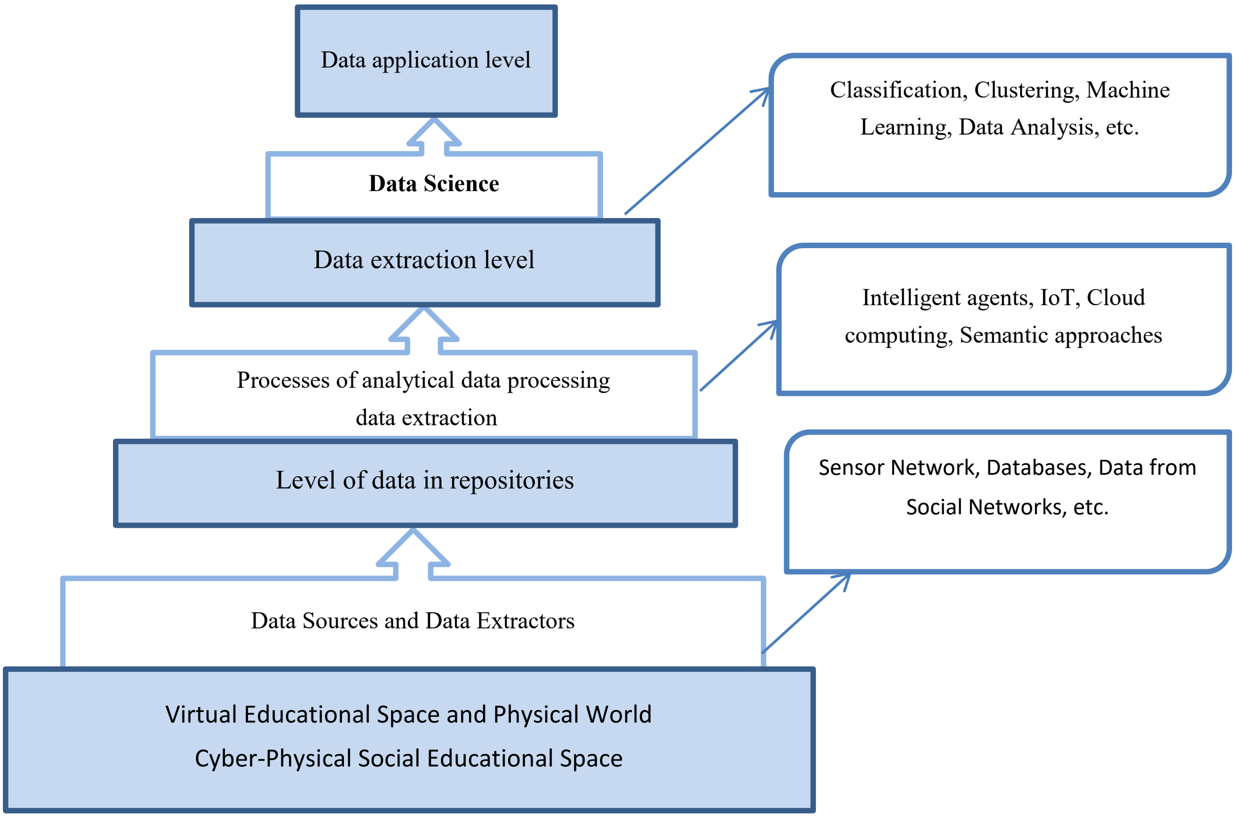

Figure 1. Functional levels in the virtual educational space

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 сиистема 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] :

- EN – *Evaluation nets* add token movement duration;
- TPN – *Temporal PNs* add transition activation moment;
- CPN – *Color PN*. In colored Petri nets, each token has a color and can only move along arcs of the same color;
- SPN – *Stochastic PN*. Arc selection in stochastic Petri nets is based on a randomly generated number;
- 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);
- PRON – *Pro-Net* introduces transition type;
- PTN – *Predicate/Transition Net* . These networks define a transition condition;
- MN - *M Net*, where the tokens also perform additional procedures;
- 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.
- 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.

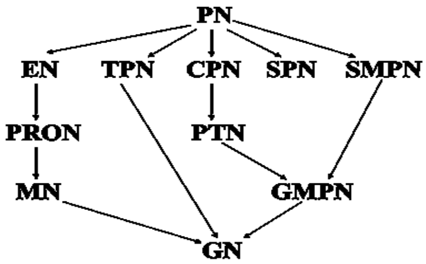

Figure 2. Extensions of Petri nets

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 incresed.

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

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

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.

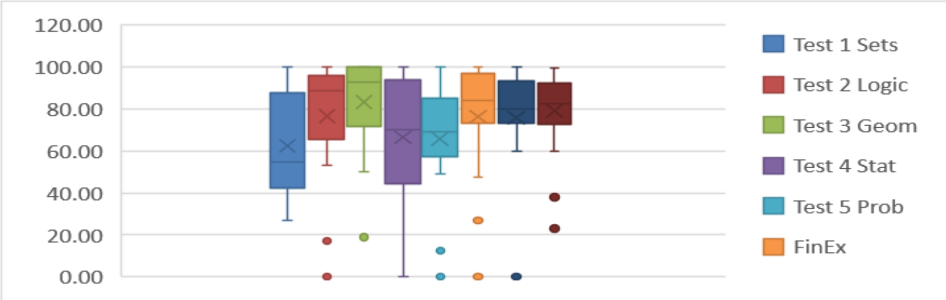

Figure 3. Graph regarding the distribution of grades by modules

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.

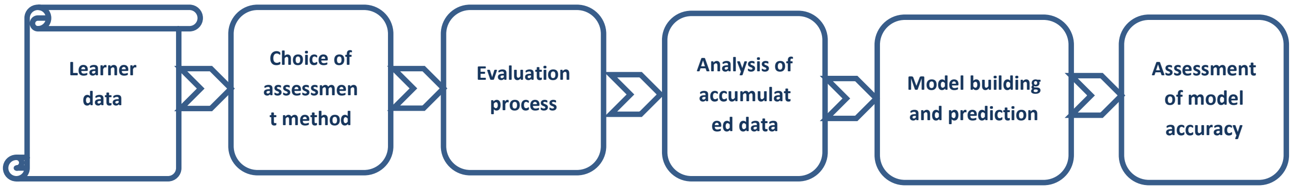

Figure 4. Multi-step process of analyzing and predicting learner knowledge

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 x 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) *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) *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) *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) *Recall = TP/(TP+FN)*

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

(7) *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) *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), qwizz 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*.

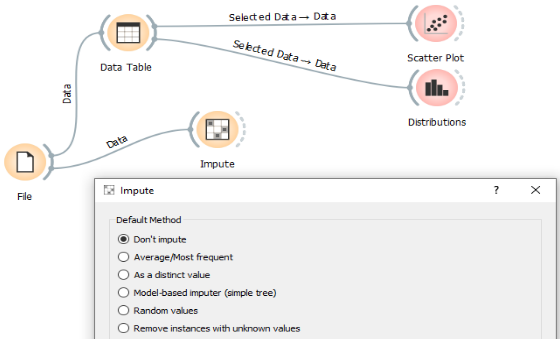

Figure 6. Applying data loading and cleaning tools

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.

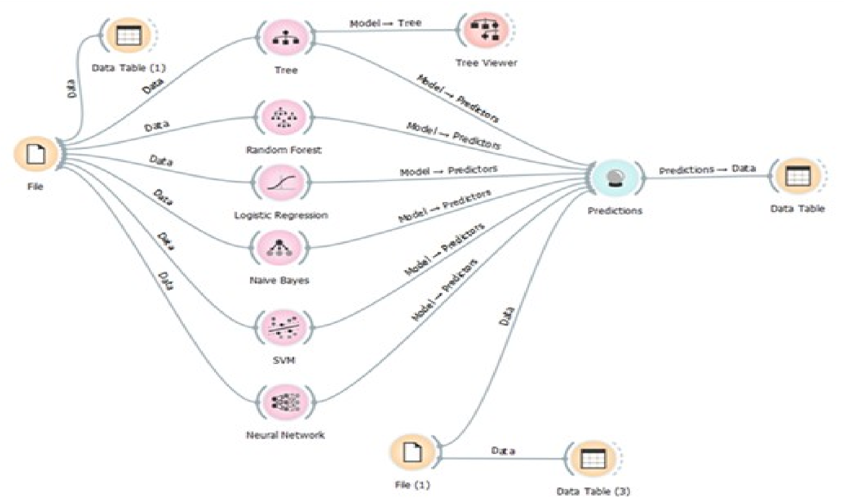

Figure 7. Model training and prediction workflow

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.

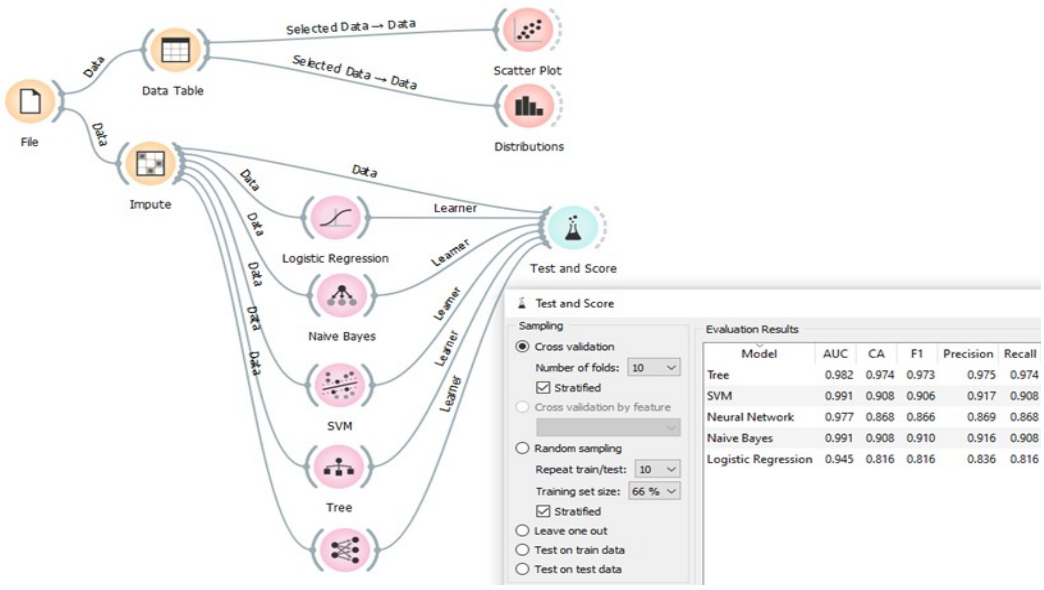

Figure 9: Scoring workflow with the " *Test and Score* " tool

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.

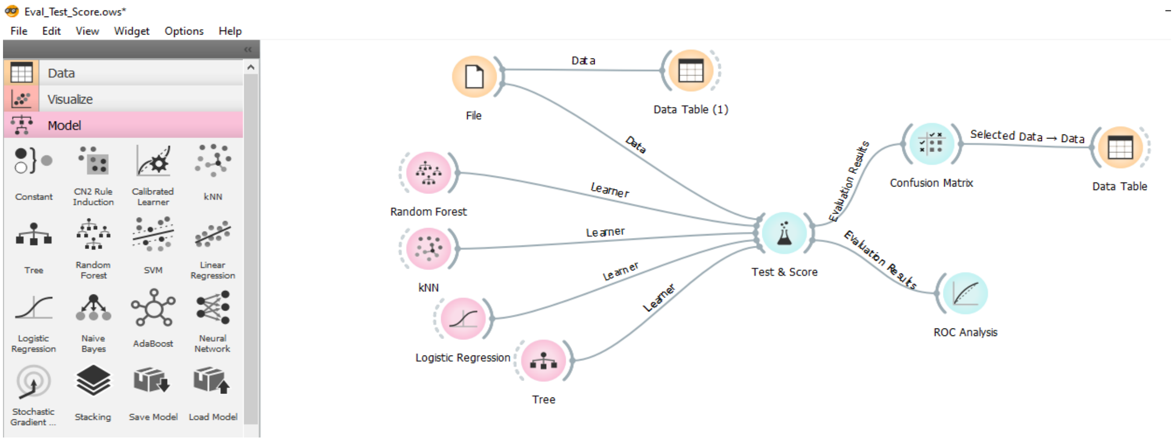

Figure 11. Model creation and evaluation workflow using *Confusion Matrix* and *ROC Analysis*

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 for Tree (showing number of instances)

		Predicted					
		Bad	Excellent	Good	Middle	Very Good	∑
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
	∑	13	22	15	7	19	76

Confusion matrix for Logistic Regression (showing number of instances)

		Predicted					
		Bad	Excellent	Good	Middle	Very Good	∑
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
	∑	15	20	19	3	19	76

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.

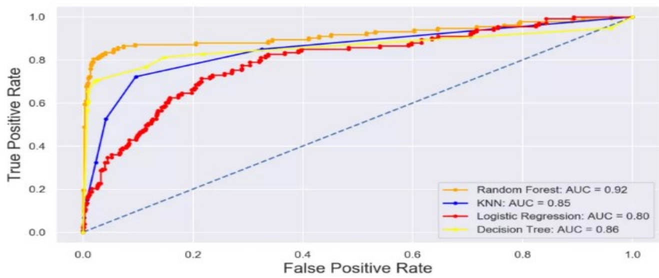

Figure 13. Visualization with the tool *ROC Analysis* in *Orange*

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/Redude 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* .

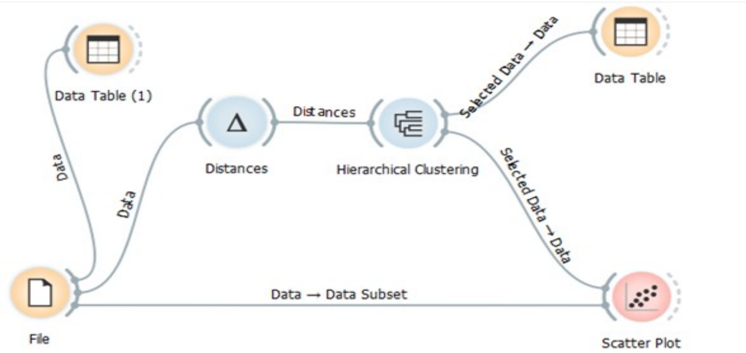

Figure 1 5 . Hierarchical clustering workflow

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;
- ***l ength 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.

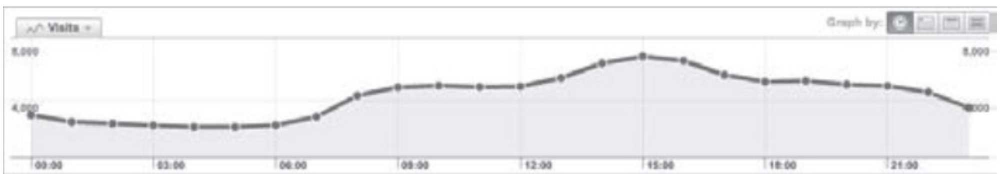

Figure 16. Graph aa hourly trend with Google Analytics.

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) $A = \{(x, \mu A(x))\}$, 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***":
  $\mu high(x) := P(M < x)$
- the indicator M will be classified as "***low***" (the negation of belonging to the "high" class):
  $\mu low (x) := 1 - \mu high (x)$
- "***medium***" classification is defined as:
  $\mu medium (x) := 1 - abs(\mu high (x) - 0.5) - abs(\mu low (x) - 0.5)$.

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:

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

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.

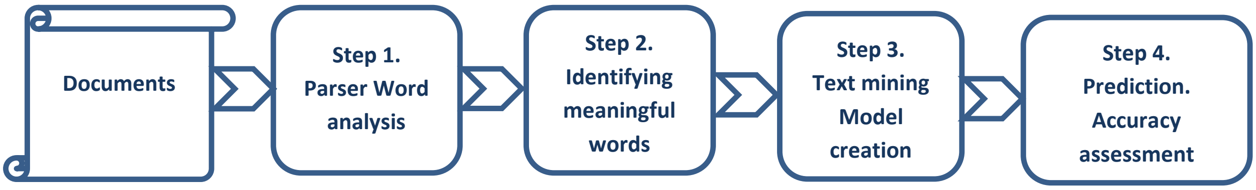

Figure 17. Multi-step process of analysis and prediction of documents from a given domain

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).

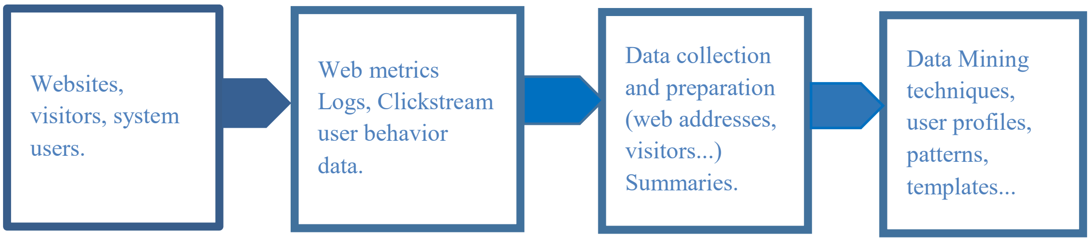

Figure 18. Web metrics data extraction architecture

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:

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

where $Mark\_Subterm[i]$ is the evaluation of the *ith* 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:

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

where $Mark\_Term[j]$ 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/).

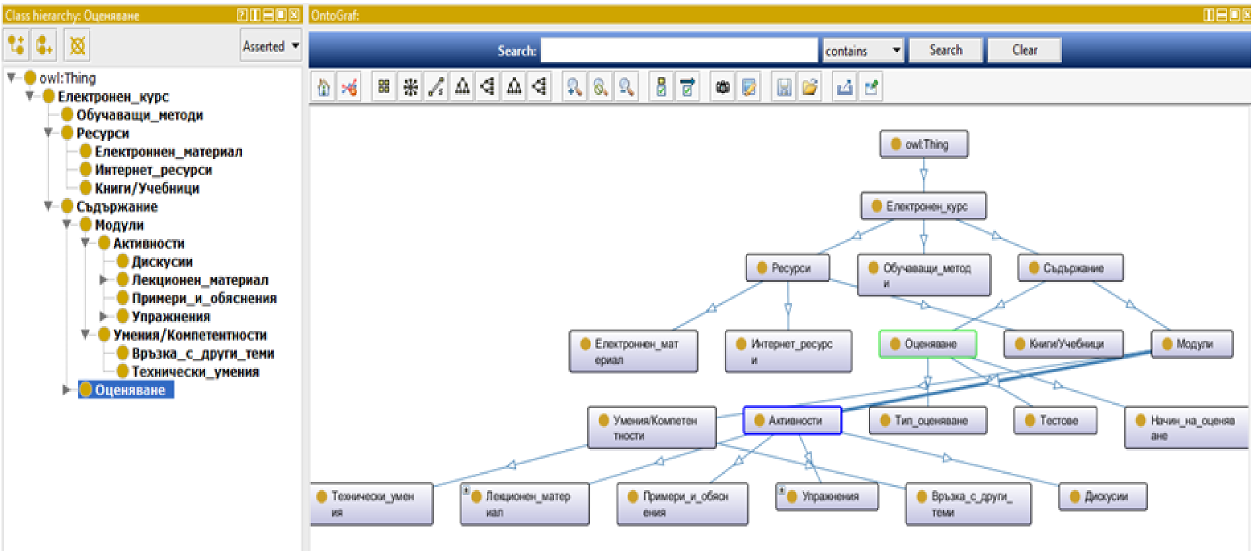

Fig. 19. Ontological model of an electronic course through the Protege system

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.

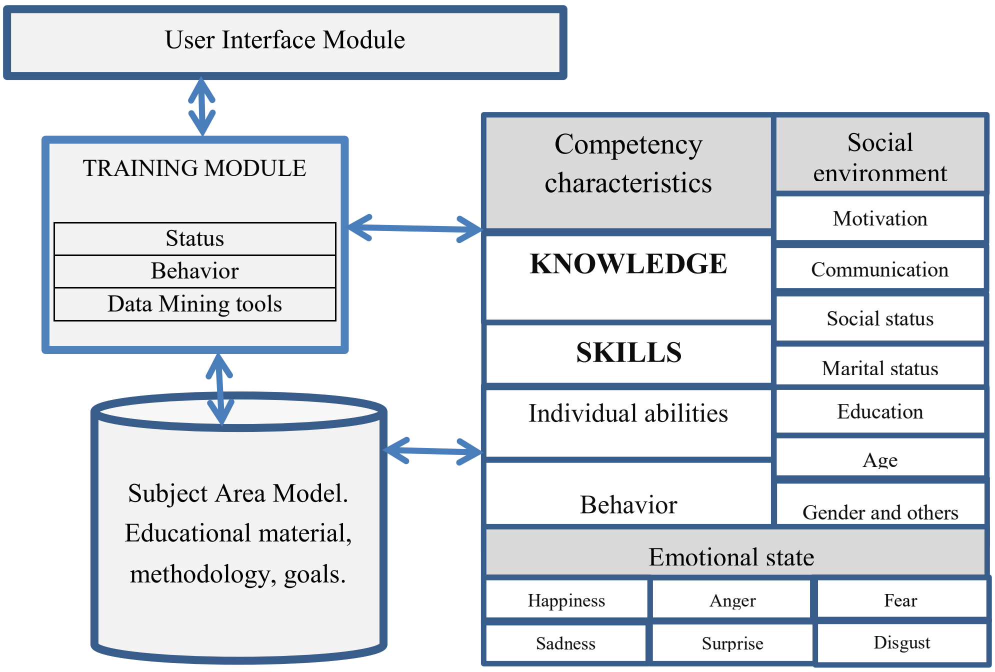

Fig. 20. Architecture of a learning system with learner model

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) $Mark\_current = 1 - Count\_error / Count$

Here, *Mark _ 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 *Count _ 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) $Mark = 2 \,.\, Mark\_current - 1$

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

(14) $Mark = 1 - 2 \,.\, Count\_error / Count$

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) $Mark\_new = (1 - \mathrm{K}) \,.\, Mark + K * \mathrm{Assessment\_of\_the\_assignment}$

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) $K = 0.5 \,.\, (Count - 1) / Count$

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) $Evaluation\_new = (1 - \mathrm{K}) \,.\, Rating\_old + K \,.\, Problem\_Assessment$

but here *problem_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) $K = Problem\_Difficulty / Unit\_Difficulty$

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.

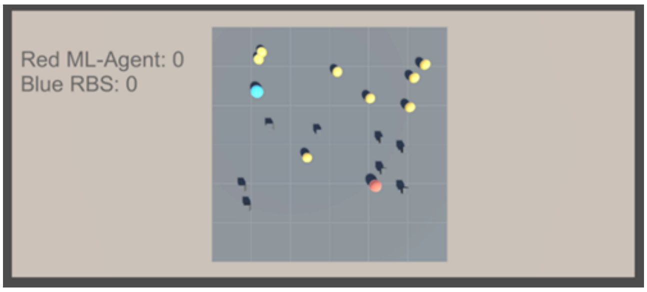

Figure 21. View of the developed software

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.

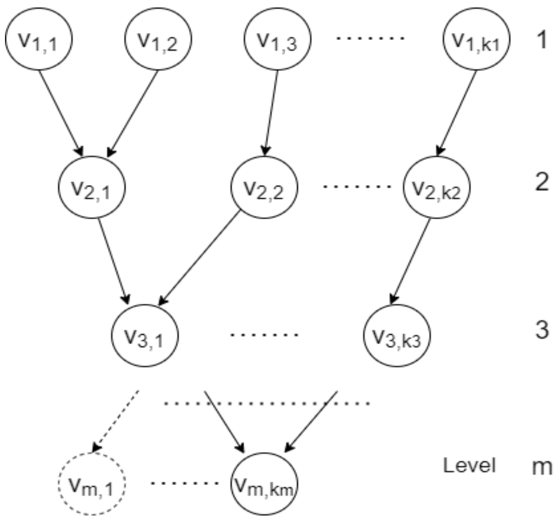

Figure 22. a) ***Model 1***: Hierarchical tree-like organization of evaluation components

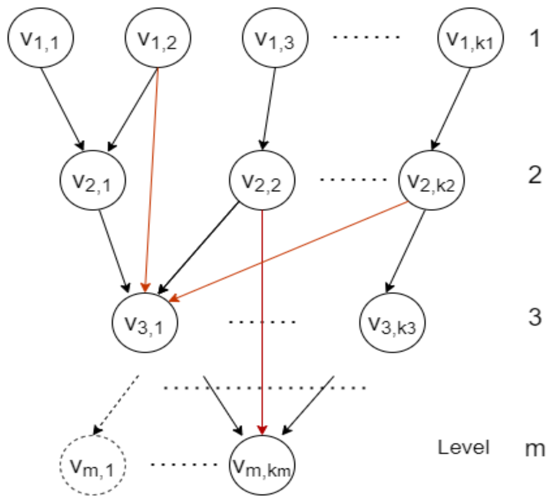

Figure 22. b) ***Model 2***: Hierarchical graph organization of evaluation components

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

Level 1 components are specific values obtained during evaluation. At each subsequent leveli $= 2 \dots m$ 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 $k_s \leq k_{s-1}, \mathrm{s} = 2, \dots, m$, 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) $v_{i,j} = f_{i,j}(P_{i-1,j}), P_{i-1,j} = \left(v_{i-1,s_1}, v_{i-1,s_2}, \dots, v_{i-1,s_j}\right) \subseteq V_{i-1},$

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

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 finalNis 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.

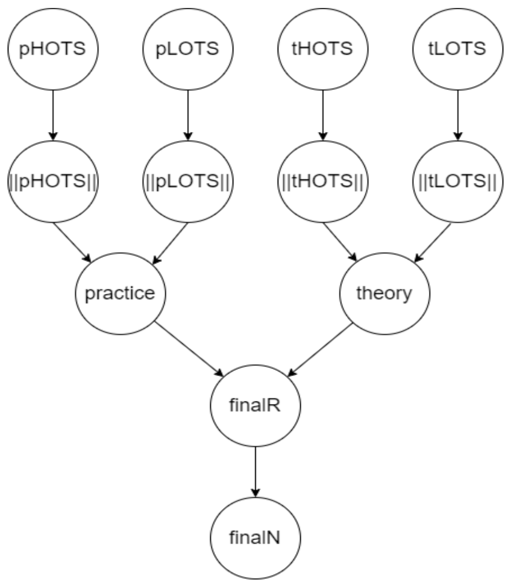

Figure 23.a) ***Option 1***: Theoretical-practical hierarchical multi-component assessment for ***Model 1***

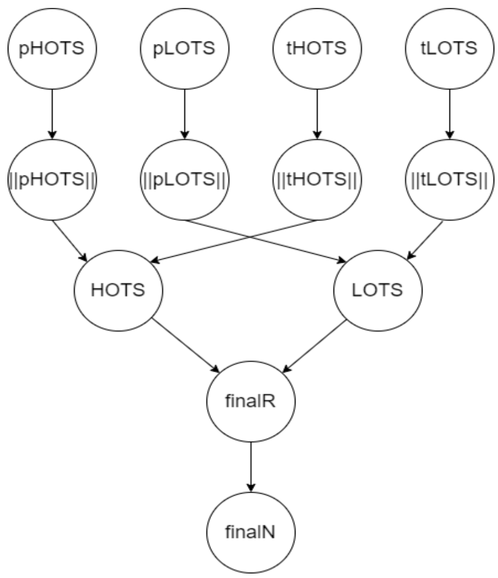

Figure 23, b) ***Option 2*** : HOTS - LOTS hierarchical multivariate estimation for ***Model 1***

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

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

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

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

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:

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

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

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

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

Represented by IF - THEN logic, the formulas look like this:

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

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

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

Arithmetic average values from the sub-components are not always a good solution and therefore, a subjective judgment is reached by the evaluator. There are many specific cases that need to be formalized to describe the logic of the estimator. Some of the considerations used in our assessment to assess the ***practice component*** are as follows:

- If pHOTS or pLOTS is 2, the score of the practice component is at most 3, even if the other sub-component has a score of 6.
- If the pHOTS score is high, and pLOTS is low, doubts arise about dishonest actions of the student when solving the task forming the pHOTS assessment. Then, we set the score to be lower than the arithmetic mean. This happens most often when pHOTS is close to 6 and pLOTS is close to 3.
- If pLOTS is high and pHOTS is low, then the knowledge shown on the test ensures that the student has the necessary foundation for future development in the discipline. Although it did not perform well in solving the task, it deserves a boost within reasonable limits. They must be pre-formalized with precise numerical values and rules.

Similar considerations apply to theoretical knowledge. That being said, high tHOTS scores are well tolerated even with the partial lack of tLOTS theoretical knowledge.

The rules used in the experiment for setting the assessment of practice and theory in ***Option 1 for theoretical-practical hierarchical multicomponent assessment*** are:

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

$$IF\ \frac{\|pHOTS\| + \|pLOTS\|}{2} < 2.5\ THEN\ practice\ IS\ 2$$
$$ELSE\ IF\ \|pLOTS\| = 2\ OR\ \|pHOTS\| = 2\ THEN\ practice\ IS\ 3$$

$$ELSE\ IF\ (\|pHOTS\| - \|pLOTS\|) \geq 2\ THEN\ practice\ IS\ \left(\frac{\|pHOTS\| + \|pLOTS\|}{2} - 0.5\right)$$

$$ELSE\ practice\ IS\ \left(\frac{\|pHOTS\| + \|pLOTS\|}{2}\right)$$

(24. 2) $\boldsymbol{theory}(\|\boldsymbol{tLOTS}\|, \|\boldsymbol{tHOTS}\|)$:

$$IF\ \frac{\|tHOTS\| + \|tLOTS\|}{2} < 2.5\ THEN\ theory\ IS\ 2$$

$$ELSE\ IF\ (\|tHOTS\| - \|tLOTS\|) \geq 2\ THEN\ theory\ IS\ \left(\frac{\|tHOTS\| + \|tLOTS\|}{2} + 0.3\right)$$

$$ELSE\ theory\ IS\ \left(\frac{\|tHOTS\| + \|tLOTS\|}{2}\right)$$

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) $\boldsymbol{finalR}(\boldsymbol{theory},\ \boldsymbol{practice})$:

$$IF\ \frac{theory + practice}{2} < 2.5\ THEN\ finalR\ IS\ 2$$

$$ELSE\ IF\ theory = 2\ OR\ practice = 2\ THEN\ finalR\ IS\ 3$$

$$ELSE\ finalR\ IS\ \left(\frac{theory + practice}{2}\right)$$

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

(26) $\boldsymbol{finalN}(\boldsymbol{finalR}\,) = ROUND(finalR)$

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.

CHAPTER 4. GENERALIZED NET MODELS IN A VIRTUAL EDUCATIONAL SPACE

4.1. A model of the processes of personalization and usage of an e-Learning environment

The purpose of the virtual educational space is to integrate the real learning process with the virtual built world in an intelligent way. Educational spaces should provide opportunities for: active and interactive participation; teamwork; searching and sharing information; discussion and presentation; producing new knowledge; supporting the activities of learners and trainers; connection with experts and last but not least personalization of training. The infrastructure of the educational space includes:

- *Building elements* of space - can be: learners, trainers, administrators, personal assistants, digital libraries, electronic services, etc.
- *Interconnections* - existing relationships between the building elements, ensuring their joint work when operating in space.

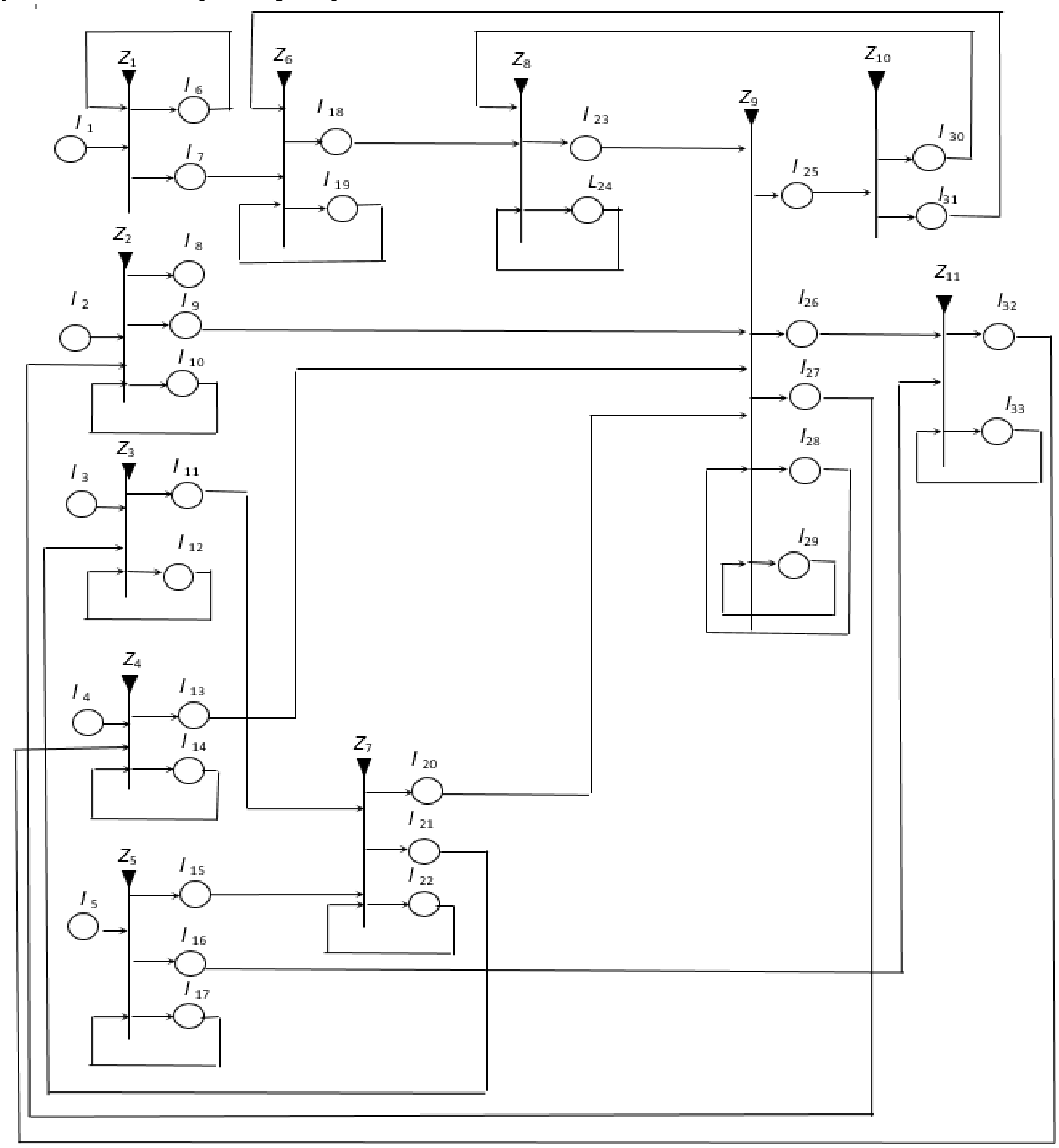

Figure 24 . GN model of the processes of personalization and usage of a learning environment

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 Υ-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*".

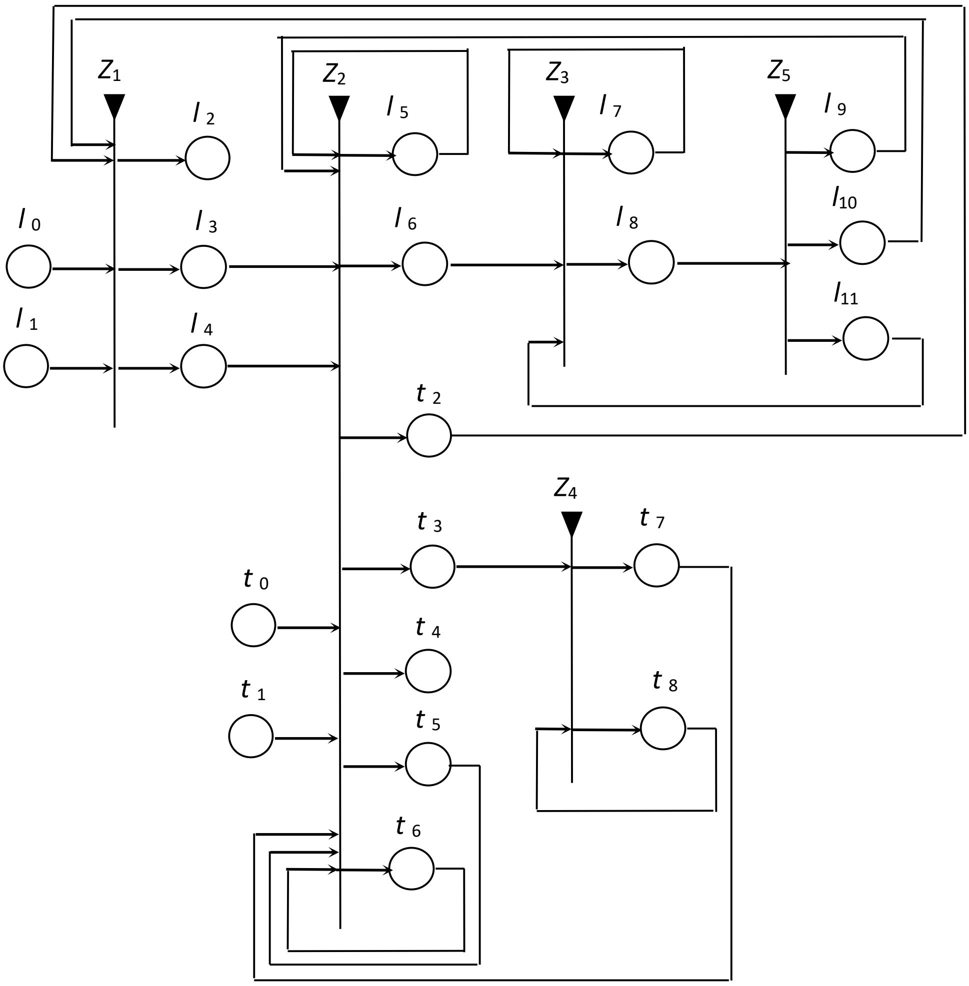

Figure 25. GN model of the process of data mining tools application in an e-Learning environments

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.

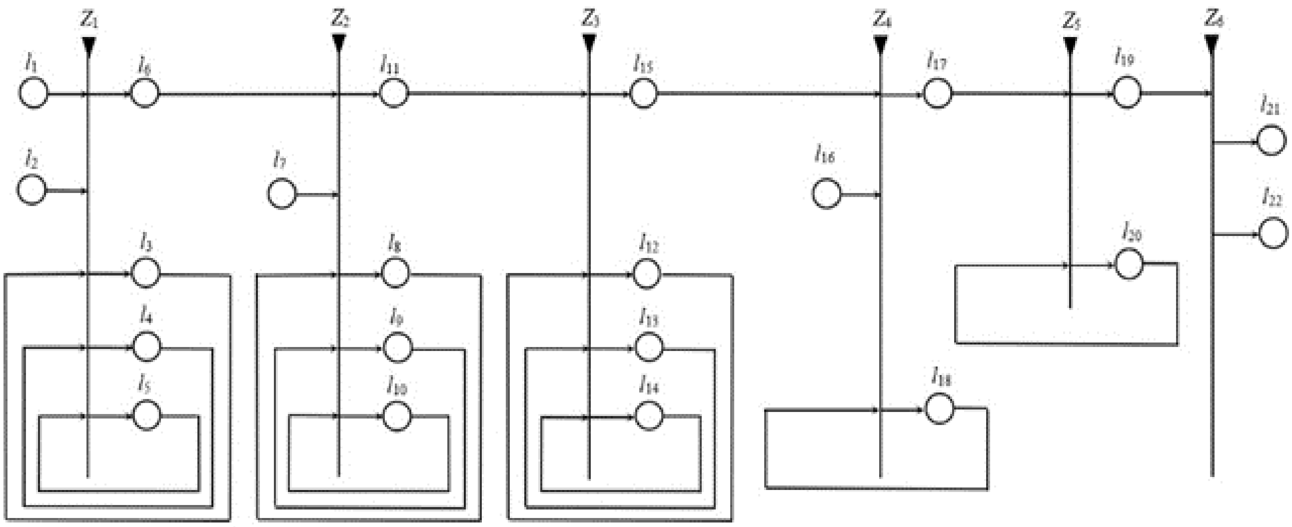

Figure 26. GN model of a multicomponent evaluation process

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.

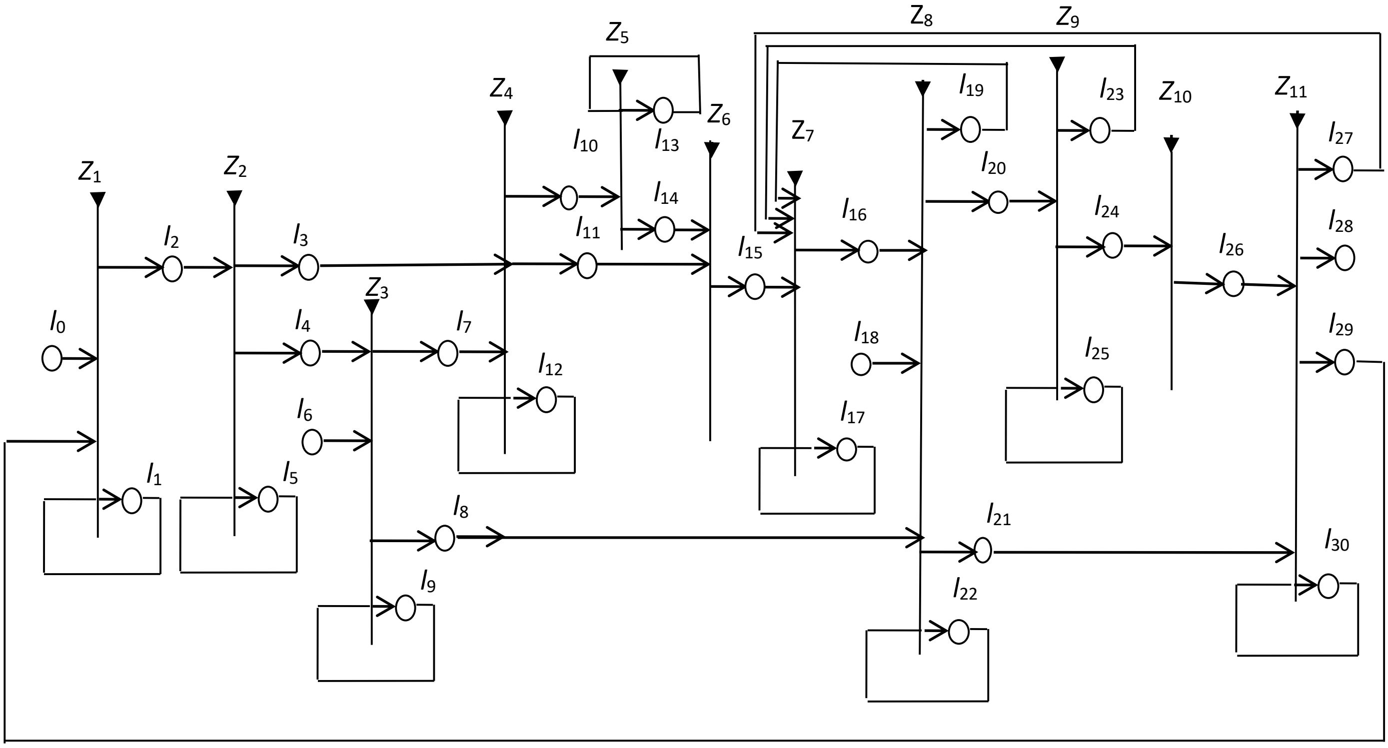

Figure 27. GN model of the processes in project-based learning

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.

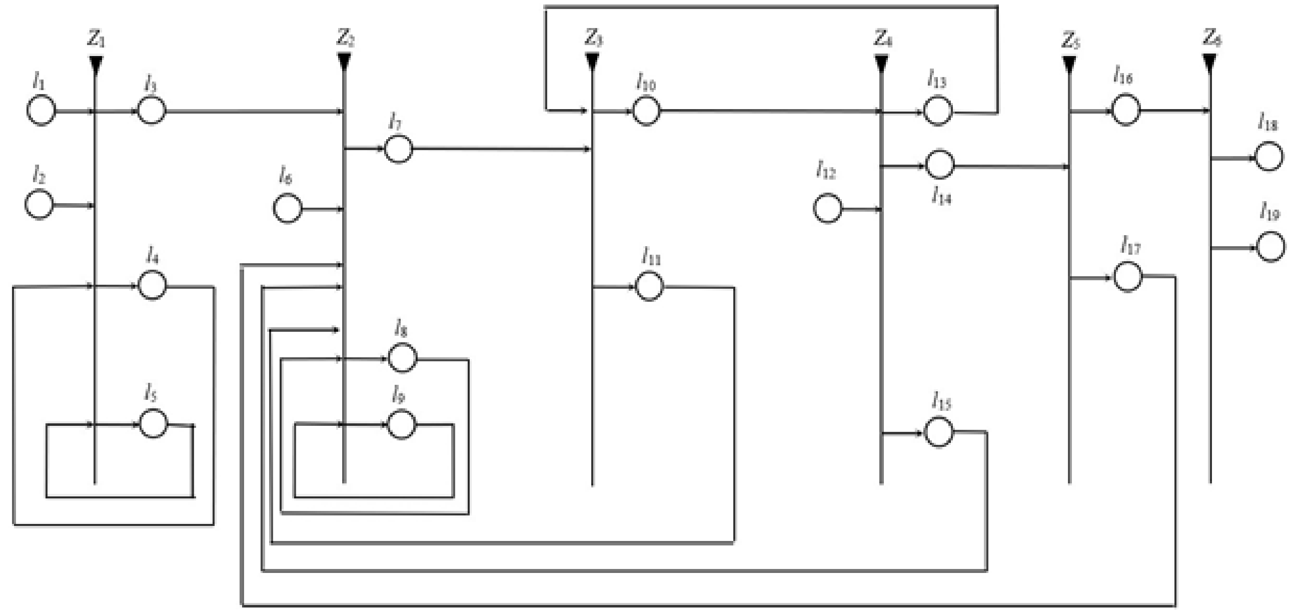

Figure 28 . A generalized net model of the gamification process of an e-learning course

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.

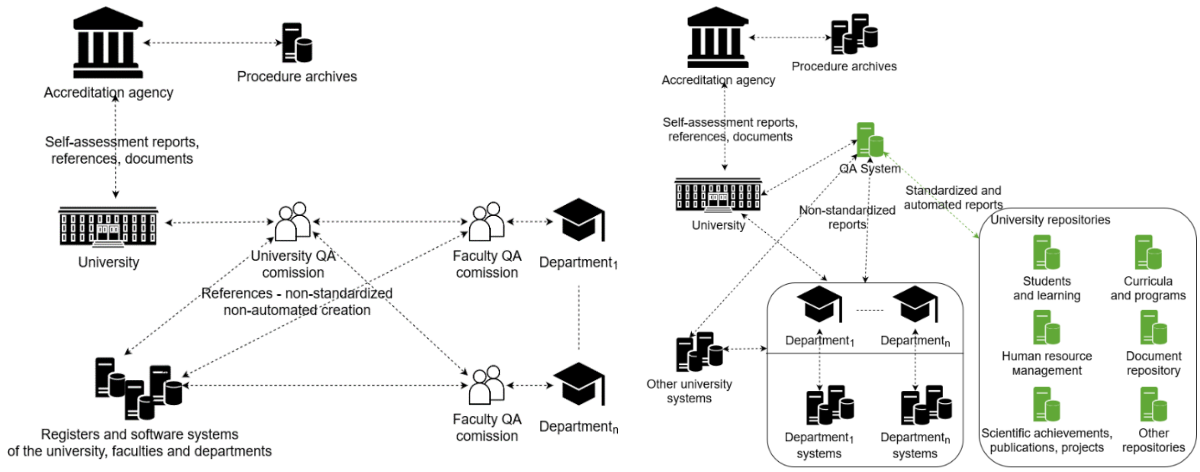

Figure 29. Model of the processes for preparing a self-assessment report by higher schools

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.

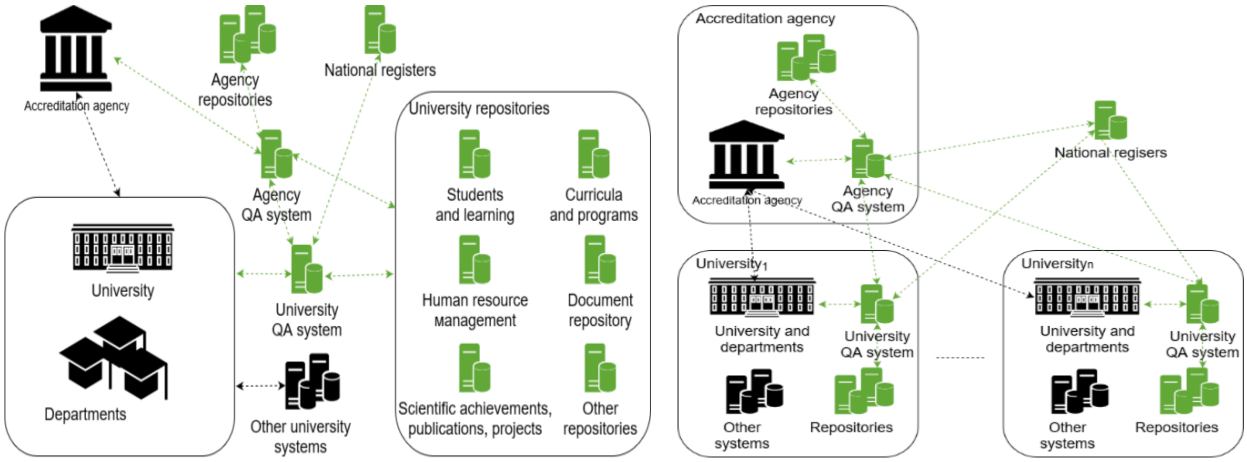

Fig. 30. A loosely centralized accreditation model

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.

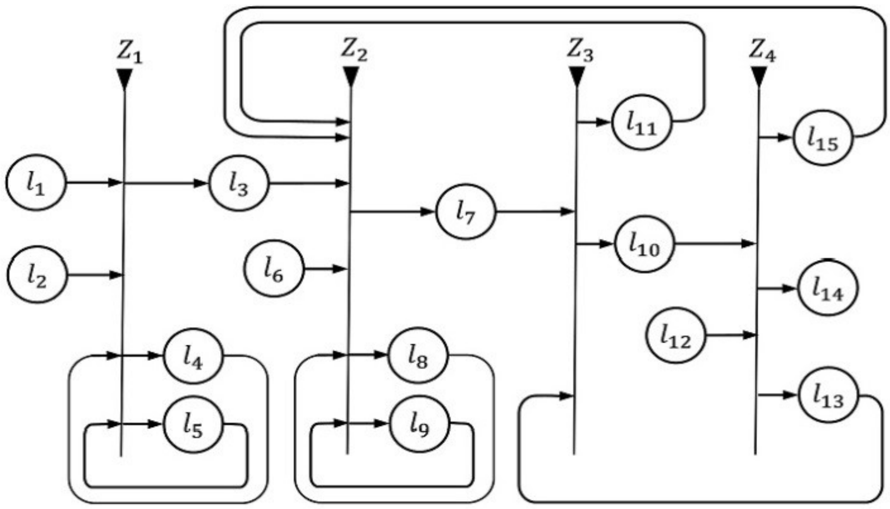

Figure 31. GN model of data processing in self-assessment in higher education institutions

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:

- o The task of *describing and summarizing data.*
- o *The segmentation (or clustering)* task.
- o *Analysis of extremes* (*outlier detection)*.
- o *Concept description* task
- o *Classification and regression*.
- o *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'Acade'mie 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'Acade'mie 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. Hazdhikoleva, 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. Magdalina 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. **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 Willey & 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. Hazdhikoleva, 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 /, 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'Acade'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'Acade'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, Schriffen 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.