INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES
BULGARIAN ACADEMY OF SCIENCES
CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **26**, No 1
Sofia • 2026 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2026-0002

A Technology for Analyzing the Transformation Process of University Students' Learning Outcomes Based on Bloom's Taxonomy

Tatyana V. Zykova, Yuliya V. Vainshtein, Mikhail V. Noskov

School of Space and Information Technology, Siberian Federal University, Krasnoyarsk, Russia
E-mails: tzykova@sfu-kras.ru yweinstein@sfu-kras.ru mnoskov@sfu-kras.ru

Abstract**:** *This paper presents a technology for analyzing the transformation of student learning outcomes (competencies) at different stages of mastering an educational program. The goal is a comprehensive approach for analyzing and predicting achievements by integrating data from curricula and the Learning Management System (LMS) Moodle digital footprints. Methods include graph analysis to model competency structures, machine learning (probabilistic matrix factorization for data imputation), Bloom's taxonomy-based classification, and multidimensional data visualization. A general research framework is presented, describing the process of processing data from various sources, including the LMS Moodle and the "Undergraduate Plans" information system, followed by data visualization, including integrated data, in the BI system Yandex DataLens. The result is a novel approach for analyzing competency development and a software module for creating an automated decision support system in the educational process, based on the analysis of learning data.*

Keywords**:** *Learning outcomes, Competency-based approach, Probabilistic Matrix Factorization, Bloom's Taxonomy, BI system, Competency ladder.*

1. Introduction

The rapid digital transformation, including in the sphere of education [1], is fostering the development of such scientific fields as educational data mining and learning analytics [2]. These two rapidly evolving areas are aimed at improving learning and educational processes. The substantial volumes of heterogeneous educational data currently accumulated in the digital educational spaces of universities [3] hold significant potential for addressing a wide range of tasks [4, 5]. Despite this, the primary sources of information about the mastery of an educational program and the achievement of learning outcomes in the form of acquired competencies typically remain data from students' intermediate and final assessments, as well as data on current academic performance and attendance. Predictive analytics models have become most widespread in scientific practice; these models use such data to forecast student success in individual courses based on various indicators [6-8]. However, this approach to assessing learning outcomes, which relies solely on current performance and temporal metrics, fails to account for the nature of how these outcomes are formed and remains largely one-dimensional. It is incapable of forecasting the quality and structure of the competencies developed, which, in turn, leads to a loss of data regarding the future graduate's professional qualifications. This is critical because a student's level of preparation reflects not only their mastery of knowledge but also their cognitive differences. Despite existing research in this area, it can be stated that one of the most discussed, contentious, and unresolved problems in education is the evaluation of educational outcomes – specifically, the search for methods and technologies to analyze the process of competency formation [9]. Currently, there is no comprehensive technological approach that accounts for the process of competency transformation. This process is dynamic and multifaceted, encompassing the acquisition, development, and modification of the knowledge, skills, and abilities students need for successful professional activity and personal growth. These competencies are formed as a result of purposeful and structured educational experience gained throughout the completion of an educational program.

When formulating a comprehensive approach to analyzing the process of competency transformation, the following models for assessing educational outcomes should be noted. For instance, the taxonomy model developed by Bloom et al. [10] is widely used to classify educational objectives by complexity levels and to specify types of learning. For relatively simple cases, such as assessing the arithmetic competencies of elementary school students, the "DINA" model was developed at the University of New Jersey based on Bloom's Taxonomy. The data for this model were presented in a matrix form, and its predictive parameters were estimated using the Expectation-Maximization (EM) algorithm for maximum likelihood estimation [11]. However, this model has not seen widespread adoption due to its high computational demands and the ambiguity in detailing the required skills. Similar models, where data is represented in matrix form, have been developed in a series of works dedicated to matrix factorization algorithms [12, 13] for building recommendation systems. In a recent study [14], a research group from Shenyang (China Northeastern University) introduced the "BloomCDM" model, which is based on the concept of matrix factorization and Bloom's cognitive theory, designed to calculate knowledge across three levels of the taxonomy. However, this result is not clearly linked to the competencies intended by the developers when designing the curricula of Educational Programs (EP). Consequently, it does not allow for the formation of a holistic understanding of students' competency profile, which is important in designing the modern educational process [15].

This paper presents an approach to analyzing the transformation of learning outcomes, using the development of mathematical competence in undergraduates from the educational program 09.03.02 "Information Systems and Technologies", specialization 09.03.02.31 "Computer Game and Application Development", as a case study. A distinctive feature of the proposed technological approach is its multi-stage process for analyzing the transformation of the Learning Outcomes (LOs) planned by the EP. This includes assessing their attainment level at various stages of instruction and their (multidimensional) visualization as a “ladder of learning outcomes”, which tracks their progression across the levels of Bloom’s Taxonomy. This framework enables the analysis of the competency transformation process by all educational stakeholders (students, instructors, university administration, and employers). A central task addressed in this study is the integration of heterogeneous educational data from two sources. The first source is the planned learning outcomes, as outlined in the EP curriculum. The second is the students’ digital footprint, generated within the university’s digital educational environment, which reflects the achieved learning outcomes for individual courses.

2. A general framework for researching the transformation of learning outcomes from curriculum to course-level results

The preparation of highly qualified graduates is a complex task dependent on numerous factors, including the quality of the educational program – which is determined by its pedagogical design, the quality of its constituent courses and practical experiences [16] – and the qualifications of the teaching staff. Furthermore, the students themselves constitute a vital component of the educational process [17].

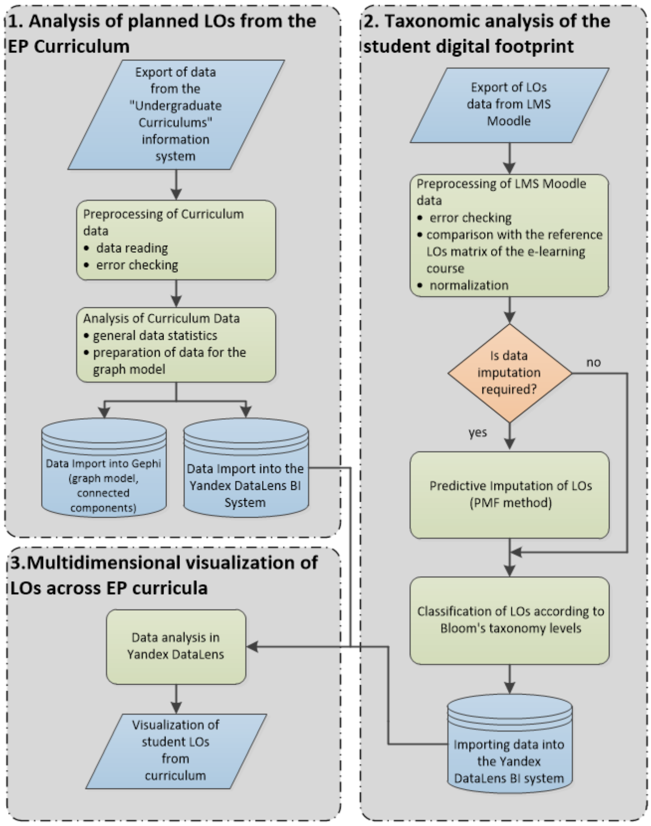

Fig. 1. General research framework

Fig. 1 presents the general research framework, which involves a three-stage process for studying the transformation of student competencies: (1) the analysis of planned learning outcomes from the educational program curriculum; (2) a component-wise taxonomic analysis of the student digital footprint from the LMS Moodle, conducted course-by-course according to the curriculum and utilizing predictive imputation for missing data; and (3) the multidimensional visualization of personal student learning outcomes in the form of a "learning outcome ladder", based on the integration and comparison of their planned and achieved results.

In the first stage of the study, an analysis of the EP curricula data is conducted using the algorithm presented by the authors in [18]. The algorithm enables the decomposition of the curriculum relative to the embedded competencies, tracks the distribution of competencies and their groups across the academic periods of the curriculum, and calculates various statistical characteristics of the curricula. These characteristics include the number of educational units in the EP and their workload (in credit units), depending on their type: courses (mandatory, stakeholder-defined, elective, optional), practical training, and the final state certification. The algorithm also calculates the relative contribution of each educational unit to the formation of competencies (EP outcomes). The algorithm implements an approach described by the authors in [19], which allows the curriculum data to be represented in a form suitable for visualization: a weighted mixed graph where the vertices represent the curriculum courses, and the edges represent interdisciplinary links, showing that the courses contribute to the formation of the same competencies.

These results serve as criteria that help to better understand the structure of the curriculum. After the analysis of the EP curriculum is completed, the results are imported and visualized in the software package Gephi (graph model) and the cloud BI system Yandex DataLens (curriculum statistical indicators). Thus, the outcome of the first stage of the study is a set of visual and tabular data that facilitates the analysis of the EP curriculum.

The second stage involves a taxonomic analysis of the student digital footprint data, exported from the LMS Moodle, for the courses of the curriculum. This data undergoes preprocessing, which includes error checking, comparison with the reference model of the course's e-learning course, and normalization to bring data from various courses into a standard format with a homogeneous structure. Subsequently, using the Probability Matrix Factorization (PMF) method, the imputation (prediction) of missing student performance data for specific curriculum courses is performed, if required. Using Bloom's Taxonomy, student educational outcomes are classified according to cognitive process levels. This procedure is carried out for individual courses in the curriculum by mapping them against the course's reference e-learning course matrix [20], which is pre-defined based on the course's learning objectives and outcomes. Following the analysis, the data on individual students' performance in the curriculum courses is imported into the Yandex DataLens BI system for subsequent visualization. The result of the second research stage is a set of visual and tabular data on student performance across various curriculum courses, reflecting their achievements across the levels of Bloom's Taxonomy.

The third stage involves the integration and comprehensive analysis of the data results from the first and second stages of the study. The transformation of learning outcomes (competencies) can be tracked through the integration of heterogeneous educational data [21], which includes data from curricula (EP), digital footprints, academic grades, and interim assessment results [22]. To achieve this, the students' learning outcomes for the curriculum courses are normalized relative to the workload defined in the EP curricula. This makes it possible to track the deviations of the achieved learning outcomes from the maximum values set by the curriculum developers for individual courses. The final result is a visualization of the transformation process of educational outcomes from semester to semester and from course to course. Based on the analysis of this visualization, stakeholders in the educational process can make managerial decisions to improve the quality of the EP and adjust the student's personal educational trajectory.

Let us now examine the outlined stages of the research in more detail.

2.1. Analysis of the planned learning outcomes in the educational program curriculum

The vast majority of educational institutions worldwide and in Russia utilize specialized automated software systems for designing and managing the components of EPs. As the starting point for this research, the Information System (IS) "Undergraduate Curricula", developed by Mathematical Modeling and Information Systems Laboratory Limited Liability Company (MMIS LLC), was selected. This system is currently one of the most widely used software complexes of its kind in Russia. Leading universities in the country, including Siberian Federal University, use the IS "Undergraduate Curricula" for developing curricula, course syllabi, planning teaching loads, and creating individual faculty plans. The system allows for data export; therefore, for a comprehensive analysis of a large number of curricula from various institutions not connected by a single database, a software solution (algorithm) was necessary. This algorithm was required to rapidly extract all data from the files, perform preprocessing, and conduct subsequent analysis. A curriculum is the core document of an EP, containing information on intended learning outcomes (i.e., competencies mapped to specific courses within the curriculum). The IS "Undergraduate Curricula" allows for the validation of curricula against key regulatory indicators, such as program duration, total workload in credit units per academic year or for the entire program, maximum student load (per week, semester, or examination session), and the number of hours allocated for contact work, among others. However, this lacks built-in tools for tracking the attainment of core educational outcomes, such as the distribution of competencies and their groups across the curriculum's components. Consequently, it is impossible to assess their individual contribution to the overall EP.

Therefore, in their previous work [18], the authors presented a software implementation of an algorithm designed to preprocess curriculum data and present it in a format suitable for analysis, identification of patterns and key characteristics of curricula and Educational Programs (EPs), as well as for comparative analysis of EPs. The proposed algorithm calculates various statistical characteristics of curricula, such as the number of educational units within an EP and their workload (in credit units), based on type:

- courses (mandatory, stakeholder-defined, elective, optional);
- practical training;
- final state certification.

The algorithm also calculates the relative contribution of each educational unit to the formation of competencies (EP outcomes). A separate software module was implemented to perform a comparative assessment of key calculated metrics across several loaded curricula. Essentially, the algorithm's output is a new set of tables (the analyzed EP curriculum data), which constitutes the educational program curriculum database (the Curriculum DB).

For subsequent analysis, the Curriculum DB was imported into the cloud BI system Yandex DataLens. This system allows for connection to various data sources, data modeling, analytical computations, building charts and visualizations, assembling dashboards, and providing collaborative access to analytics. DataLens is registered in the domestic software registry and complies with both Russian and international standards for security and the processing of sensitive data.

Various visualizations were developed in Yandex DataLens to track and analyze key metrics of EP curricula, both for individual programs and across the entire set. The Russian educational system uses a competency-based model categorized into three main types: Universal Competencies (UC) – transversal skills such as communication and critical thinking; General Professional Competencies (GPC) – foundational skills within a specific field; and Professional Competencies (PC) – specialized skills for particular job functions. Furthermore, some curricula also include General Learning Competencies (GLC). The Federal State Educational Standard (FSES) includes a numbered list of competencies from these groups. Curriculum developers use this list when designing EP by incorporating various competencies into the plan. Fig. 2 provides an example of a visualization from the Curriculum DB, showing the distribution of workload allocated in the curriculum to all UC, GPC, and PC.

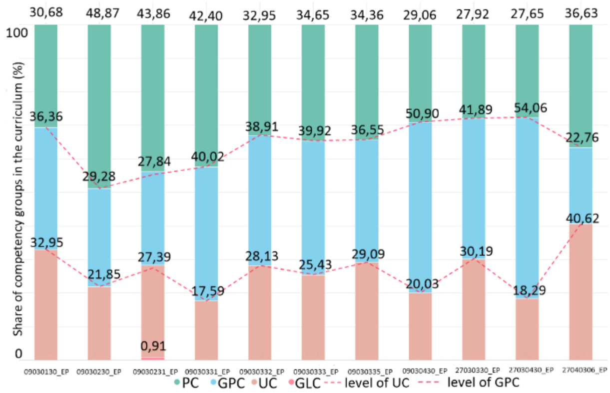

Fig. 2. Distribution of workload for UC, GPC, and PC in the curriculum

A curriculum allocates a varying number of hours and credit units for the implementation of its courses, and typically assigns no more than three competencies to each. The methodology for calculating the workload of competencies and their relative weight was described in detail in our previous work [18]. For instance, the curriculum of the educational program 09.03.02 "Information Systems and Technologies", with the specialization track 09.03.02.31 "Computer Game and Application Development", uses a treemap diagram (Fig. 3) to visually represent the list of competencies (and their relative weight) incorporated by the developers into the educational program during its design.

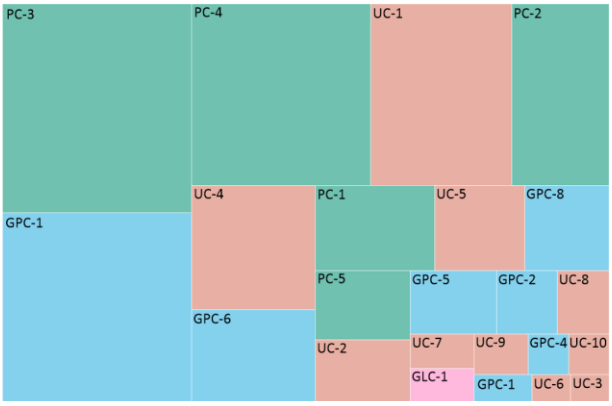

Fig. 3. Competencies and their relative weight in curriculum 09.03.02.31

The presented plan includes various competencies, such as:

- GPC-1. The ability to apply knowledge in natural sciences and general engineering, methods of mathematical analysis and modeling, and theoretical and experimental research in professional activities.
- GPC-2. The ability to understand the principles of operation of modern information technologies and software, including domestic products, and to use them in solving professional tasks.
- UC-1. The ability to search, critically analyze, and synthesize information, and apply a systems approach to problem-solving, etc.

The diagram automatically generates rectangles representing each competency. This visualization helps to intuitively understand the contribution of each competency to the educational program, as the size of a rectangle corresponds to the workload allocated to that competency.

Next, we can separately examine the competencies presented in the curriculum. For example, the formation of competency GPC-1 in curriculum 09.03.02.31 is allocated 32.75 Credit Units (CU). The appearance of fractions in the calculated indicator is because each course in the curriculum typically contributes to the formation of one to three competencies, and when calculating the workload for any one of them, an assumption is made about the equal contribution of these competencies to the course. A detailed picture of the formation of GPC-1 in curriculum 09.03.02.31 is presented in the ring diagram (Fig. 4). It shows all the courses in the curriculum that are directly involved in its formation, as well as the workloads (in CU) allocated to GPC-1.

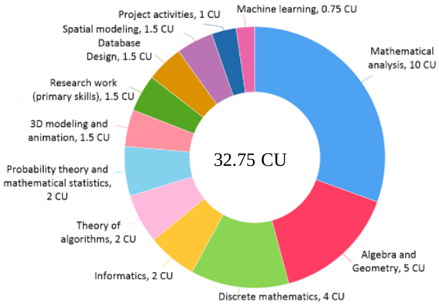

Fig. 4. Distribution of CU for the formation of GPC-1 in curriculum 09.03.02.31

For the convenience of analysis, let us represent the curriculum as a mixed graph $G = (V, E, A)$, where V is the set of vertices defined by the list of courses included in the curriculum, $E \subset \{uv: u, v \in V\}$ is the set of undirected edges, and $A \subset \{\overleftarrow{uv}: u, v \in V\}$ is the set of directed edges (arcs), where u is the tail and v is the head of the arc. Let N denote the total number of courses implemented in the curriculum, and $\{C_i\}$, *i*=1, 2, …, *N*, the set of curriculum competencies. Each course is a vertex. v of the graph – will be assigned a set of competencies $\{C\}_v$ allocated to that course. An edge in a graph G is understood as an *interdisciplinary connection* defined by the intersection of the sets $\{C\}_u$ and $\{C\}_v$, which occurs when two courses in the curriculum contribute to the formation of one or several identical competencies. An edge exists if $\{C\}_v \cap \{C\}_u \neq \varnothing$. An edge uv in the graph will be undirected if the corresponding curriculum courses (graph vertices) are delivered simultaneously, for example, in the same academic semester or year. A directed edge $\overleftarrow{uv}$ is considered such when the curriculum courses (graph vertices) have temporal priorities, meaning the courses must be delivered in different time periods. The weight w_{vu} of an edge can be defined, for instance, as the arithmetic mean of the CU corresponding to the vertices of the edge [19]. Fig. 5 shows a fragment of the mixed graph. G representing curriculum 09.03.02.31, visualized using the open-source network analysis software Gephi 0.10.1, which is developed on the Java NetBeans platform.

To better understand the graph structure, let us consider an example. During curriculum design, competencies are distributed among courses based on their content alignment: each competency is assigned to those courses within which the necessary knowledge, skills, and abilities for its development are most organically and comprehensively acquired. The principle of balanced student workload is also considered to ensure the feasibility of objective assessment and the attainability of the declared set of competencies for each course. Consequently, a course is typically assigned between one and three competencies.

For the course *Mathematical Analysis*, developers assigned competency GPC-1; for *Probability Theory and Mathematical Statistics*, they also assigned GPC-1; and for the course *Informatics*, they assigned GPC-1 and UC-1. Since the courses *Mathematical Analysis*, *Probability Theory*, *and Mathematical Statistics* share the common competency GPC-1, the graph vertices corresponding to these courses will be connected by a common edge. This edge will be directed, as the courses are taught in different academic years. The graph vertex corresponding to the *Informatics* course will also be connected to them, but the edge between it and *Mathematical Analysis* will be undirected (as these subjects are studied in the first year, i.e., concurrently).

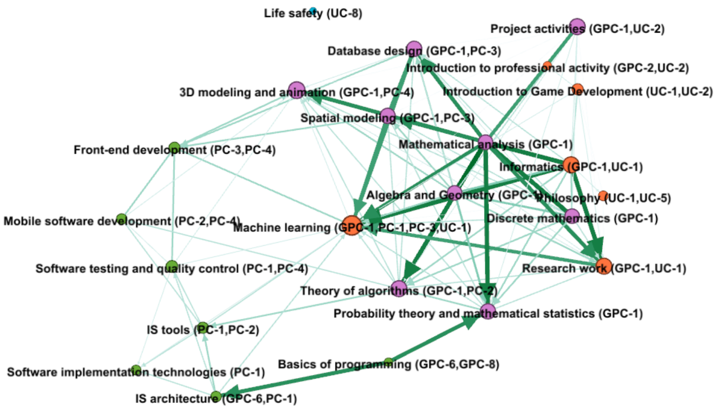

Fig. 5. Fragment of graph G for curriculum 09.03.02.31

In the mixed graph Gone can identify connected components as a set of vertices $v_0, \ldots, v_k$ and edges such that all vertices in this set participate in the formation of the same competency C_i, *i*=1, 2, …, *N*, Such a component will define *the process of forming competency*. C_i, visually indicate the curriculum disciplines involved in its formation, as well as the related competencies implemented in the disciplines included in the component. With proper visualization of the graph model (its connected components for C_i one can clearly identify the temporal stages of mastering disciplines, clusters of related disciplines, and competencies that determine the interaction of competency C_i with other components of the curriculum.

To render a graph G (its connected components), we applied the following sequence of actions in Gephi:

- Application of the Force Atlas 3D plugin. This plugin was developed on top of the two-dimensional “Force Atlas 2” and was introduced in early 2023. Visualizing a 2D layout in 3D space helps to avoid excessive node overlap, allows for a more precise understanding of the patterns and structure of the object under study, and better reveals the centrality of the graph node with the maximum set of connections to other nodes. The 3D capability enables the visualization of more complex cluster structures by arranging nodes on parallel planes.
- Configuration of the graph model, which included ranking nodes by degree, rendering edges according to their weights, and utilizing the properties of modularity and clustering.
- Application of the Network Splitter 3D plugin. It allows for the separation of the graph layout into individual layers, which are clusters defined by a specific

variable; these layers can be used in ranking and/or splitting procedures. Network Splitter 3D calculates and segments levels relative to values within any ranges.

Fig. 6 shows an example of a connected component of a graph G, illustrating the process of forming competency GPC-1 for curriculum 09.03.02.31. It clearly displays all curriculum disciplines involved in the formation of GPC-1. The nodes of the component are programmatically separated into distinct layers (the semesters/courses in which the corresponding disciplines are studied).

This visual representation as a connected component aids in better understanding the structure of how the corresponding competency (the educational outcome) is formed. The presence of clearly visualized interdisciplinary connections enables the analysis of the curriculum at the level of individual course syllabi and allows for a more nuanced understanding of knowledge continuity from the perspective of achieving educational outcomes. The proposed approach facilitates curriculum analysis, revealing its characteristics, which help in assessing the quality of the implemented EP, its pedagogical design strengths and weaknesses, and enables comparative analysis of different EPs to identify successful design solutions for improving the quality of designing new programs or modernizing existing ones. Furthermore, the result of the first research stage is a pathway that a student must follow to develop the curriculum's competency during their progression through the EP.

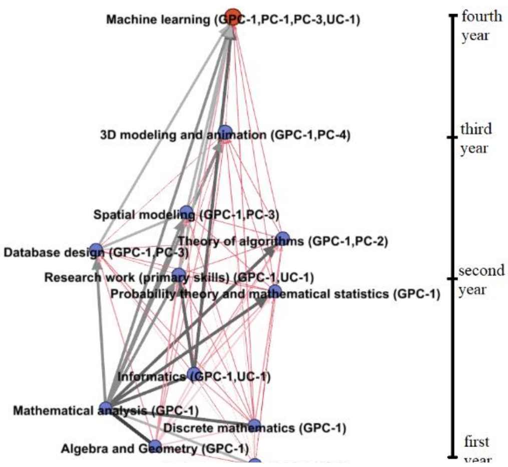

Fig. 6. Connected component of graph G, demonstrating the formation process of GPC-1 for curriculum 09.03.02.31

2.2. Taxonomic analysis of student digital footprint data across curriculum disciplines in an educational program

At the second research stage, an analysis of student digital footprint data for individual curriculum disciplines of the educational program was conducted. The LMS Moodle served as the data source. Moodle allows for the export of electronic gradebook data for individual disciplines in a tabular format (files in .xlsx, .csv). At the initial stage, to improve the accuracy and reliability of the results, it is necessary to preprocess the data from the discipline's E-Learning Course (ELC). This preprocessing includes the following steps:

1. Error control. Since creating a course in Moodle involves the comprehensive configuration of course elements with specified assessment formats and grading scales, error control does not require excessive time. It is necessary to verify data format compliance. Additionally, an analysis of student activity logs can be performed to identify potential anomalies in their behavior, such as multiple login attempts, etc.

2. Benchmarking against the ELC standard. This stage includes:

- content analysis: reviewing the course content for compliance with the curriculum's course syllabus and the standards established by the educational program;
- performance (engagement) analysis: checking grades and activity against the course's standard indicators (number of completed modules, time taken to complete assignments, etc.);
- contextual analysis: evaluating the data for alignment with the specific conditions and peculiarities of each course, considering that different courses may have different success criteria.

3. Data enrichment. Adding supplementary variables that may influence the analysis, for example, data on student performance and attendance in the discipline's in-person classes.

4. Data normalization. This stage involves standardizing the data to a common format to enhance comparability for subsequent work and entails:

- scaling numerical data: transforming data to a unified scale;
- data merging and grouping: combining data from different sources and grouping it according to specific criteria (e.g., by course modules or weeks of study) to improve interpretability and analysis;
- identification and removal of irrelevant data: determining variables that are insignificant for the analysis and removing them to prevent distortions in further research.

Next, we denote the group of students enrolled in the course as $U = \{u_i\}$, i = 1, 2, …, N, and the set of elements of the ELC as $V = \{v_j\}$, j = 1, 2, …, M. Thus, for each discipline, we obtain a data table (a preprocessed ELC gradebook), represented by an $N \times M$ matrix $R = (r_{ij})$. The matrix contains empty entries for some students (white squares, Fig. 7), indicating that a student did not complete the corresponding tasks in the ELC. To recover (predict) these missing values, we will use the Probabilistic Matrix Factorization (PMF) method.

Factorization methods [23], particularly PMF, represent a key approach within collaborative filtering models for recommender systems. These methods have been significantly developed, especially since the 2000 s [24, 12]. The core idea involves using matrix factorization to uncover latent factors that describe user preferences and the characteristics of associated items. These algorithms perform well on highly sparse and imbalanced datasets, where not all users have provided ratings for the entire set of items, or where these ratings are very unevenly distributed. The PMF method and its variations have also found application in the analysis of educational data [13, 14].

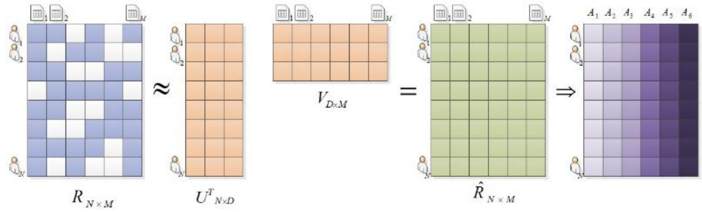

Fig. 7. Diagram of PMF application with subsequent data classification according to Bloom's taxonomy levels

The objective of the factorization method is to find (predict) the rating that a user u_i has not assigned to an item, v_j, which can nonetheless be inferred based on the ratings from other users similar to u_i characteristics or preferences. Therefore, it is necessary to find a way to fill in the missing data in the matrix inn$R$ to generate recommendations. The target $i.r_{ij}$ is the resultof the interaction between user u_i and item v_j. The tining set is defined as $\{(u_i, v_j): \exists r_{ij}, u_i \in U, v_j \in V\}$. The task is to find (recover) a dependency in the form of a function:

$$f(u_i, v_j) = \hat{r}_{ij} \approx r_{ij}.$$

Once such a function $f(u_i, v_j)$ is found, it can be used to make predictions for users and the items in question.

Following [23], our objective is to approximate the matrix R using two lower-rank matrices:

$$R_{N\times M} \approx U^{\mathrm{T}}_{N\times D} \cdot V_{D\times M}.$$

Here, each of the M features in the original matrix R is represented as a linear combination of D latent parameters. The original features – user ratings – are expressed linearly through these latent factors. The dimensionality of the matrices determines the number of latent parameters or components extracted from the data. A lower dimensionality leads to more compact data representations but may result in information loss. A higher dimensionality yields more detailed representations but increases the risk of model overfitting. Selecting the optimal dimensionality is a critical step in the matrix factorization method and depends on the specific task and dataset. Since matrices U and V have a lower rank than the original matrix R, this enables scaling the PMF method to larger datasets.

The task reduces to finding the matrices U^{T} and V, which will serve as the model parameters; we denote the set of parameters as $\theta = \{U, V\}$. The idea behind the PMF method is based on a probabilistic machine learning algorithm that relies on Bayes' theorem. According to [12], we define this model as a linear model with Gaussian observation noise. The conditional distribution over the observed ratings is defined as:

$$p(R|U, V, \sigma^2) = \prod_{i=1}^{N} \prod_{j=1}^{M} \left[\mathcal{N}(R_{ij}|U_i^{\mathrm{T}} V_j, \sigma^2)\right]^{I_{ij}},$$

where: $\mathcal{N}(x|\mu,\sigma^2)$ is the Gaussian probability density function with mean μ and variance σ^2; I_{ij} is an indicator function equal to 1 if the user i rated item j, and 0 otherwise. This formulation assumes that the random variables (user ratings of items) are independent and identically distributed; the probability of a product of independent events equals the product of their individual probabilities. This distribution is Gaussian with parameters: mean $U_i^{\mathrm{T}}V_j$, and variance σ^2.

As distributions for users and items, we adopt zero-mean spherical Gaussians:

$$p(U|\sigma_U^2) = \prod_{i=1}^{N} \mathcal{N}(U_i|0,\sigma_U^2), \qquad p(V|\sigma_V^2) = \prod_{j=1}^{M} \mathcal{N}(V_j|0,\sigma_V^2).$$

The spherical Gaussian accounts for spatial dependence in the data, assuming that data points closer to each other have a higher probability of having similar values than those further apart.

Assuming that the user feature vectors are in a matrix U are independently and identically distributed, as are the item feature vectors in the matrix V, the posterior distributions of the feature matrices U and V can be calculated using Bayes' theorem as follows:

(1) $p(U,V|R,\sigma^2) \prod_{i=1}^{N} \prod_{j=1}^{M} \left[\mathcal{N}\left(R_{ij}\middle|U_i^{\mathrm{T}}V_j,\sigma^2\right)\right]^{I_{ij}} \prod_{i=1}^{N} \mathcal{N}(U_i|0,\sigma_U^2) \prod_{j=1}^{M} \mathcal{N}(V_j|0,\sigma_V^2)$.

We will find the parameter estimates for U and V using the maximum likelihood method. This procedure is equivalent to minimizing the negative logarithm of (1). Thus, when the hyperparameters σ^2, σ_U^2, σ_V^2 are fixed, the objective function can be derived as follows:

$$E = \frac{1}{2}\sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij}\left(R_{ij} - U_i^{\mathrm{T}}V_j\right)^2 + \frac{\lambda_U}{2}\sum_{i=1}^{N} ||U_i||_{\mathrm{Fro}}^2 + \frac{\lambda_V}{2}\sum_{i=1}^{M} ||V_j||_{\mathrm{Fro}}^2,$$

where $\lambda_U = \frac{\sigma^2}{\sigma_U^2}$, $\lambda_V = \frac{\sigma^2}{\sigma_V^2}$, and $||\cdot||_{\mathrm{Fro}}^2$ denotes the squared Frobenius norm.

Subsequently, numerous well-established methods can be applied to find the parameters during model training, such as gradient-based optimization techniques. Alternatively, probabilistic modeling approaches can be utilized, including the EM Algorithm [25], Markov Chain Monte Carlo (MCMC) methods [26], and other Maximum Likelihood Estimation (MLE) [27] techniques for estimating parameters U and V [14].

Thus, by applying the PMF method to the matrix R, we obtain a matrix $\hat{R}$ with imputed data of student grades for elements of the ELC. Next, following the schema presented in Fig. 7, we perform classification according to the six levels of Bloom's taxonomy. This requires mapping the course results contained within the ELC against a benchmark ELC matrix. Such a benchmark matrix must be pre-developed for each curriculum discipline and its corresponding ELC, based on the intended learning objectives and outcomes. Technically, the procedure for creating the benchmark matrix involves tagging ELC elements according to Bloom's Taxonomy levels, specifying the points attainable for completing the corresponding task [20]. Ultimately, this process yields data on student learning across various curriculum disciplines, reflecting their achievements according to the levels of Bloom's Taxonomy.

In this framework, the reference matrix functions as a detailed blueprint that links each assignment to a specific cognitive level within Bloom's Taxonomy. All learning tasks are labeled according to the six levels: from the basic Remember and Understand, through Apply and Analyze, to the complex Evaluate and Create.

The total points a student can earn for tasks at a given level are then compared to the course's overall maximum score. This makes it possible to calculate the percentage share that the development of skills at each cognitive level occupies within the total course structure. Ultimately, by comparing actual student results with this reference distribution, one can assess not just overall performance but also understand the qualitative contribution of the course to the development of different stages of thinking: how much effort is directed towards knowledge reproduction, and how much towards its creative transformation and the achievement of the taxonomy's higher levels. Fig. 8 presents this reference distribution for the following courses: *Mathematical Analysis*, *Algebra and Geometry*, and *Discrete Mathematics*. Undoubtedly, this distribution across taxonomy levels is not fixed; it is shaped by the course materials as well as the specific learning goals and objectives of the discipline.

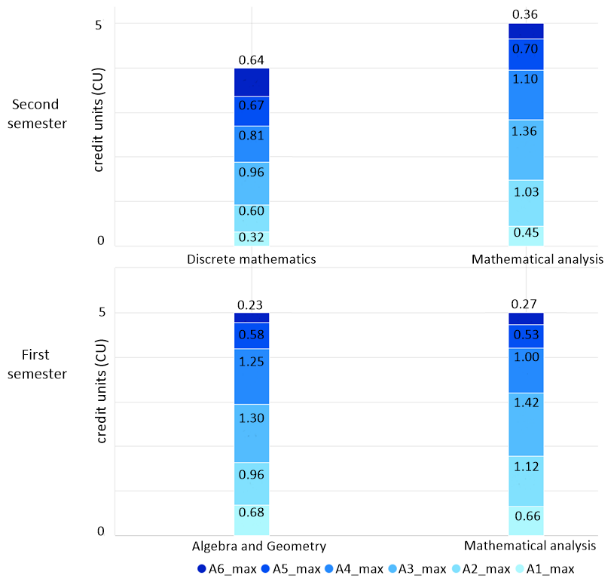

Fig. 8. Maximum Bloom's Taxonomy level indicators by discipline, normalized relative to curriculum workload

2.3. Multidimensional visualization of student learning outcomes across curriculum disciplines in an educational program

At the third stage of the analysis of educational outcomes, the data obtained in the first and second research stages are integrated and subjected to a comprehensive analysis in the Yandex DataLens BI system.

To integrate the data at the third stage, a normalization procedure was performed. Its objective was to bring the maximum (reference) Bloom's Taxonomy level indicators to a comparable form, accounting for the different workloads of the courses as stipulated by the curriculum of the educational program. Technically, this means that the calculated percentage shares for each course were adjusted based on the credit units (workload) allocated for its completion. As a result, we transition from assessing a level's share within an individual course to determining its specific weight within the overall structure of competency mastery across the entire educational program. For example, Fig. 8 presents the result of this operation for competency GPC-1 from curriculum 09.03.02.31, broken down by the mathematical disciplines studied in the first year of the program.

Consequently, the third stage involves comparing the results achieved by students across various disciplines within the curriculum against the specified maximum benchmarks defined for those disciplines in the curriculum implementation. This enables a more comprehensive assessment of how the knowledge and skills acquired from individual subjects contribute to the broader goal of achieving educational outcomes (competencies). This approach not only facilitates the evaluation of student achievement but also allows for necessary adjustments to the educational process, ensuring its alignment with modern requirements and standards. Fig. 9 provides an example of integrating the learning outcomes achieved by students A, B, and C – categorized according to Bloom's Taxonomy levels – compared against the planned outcomes for the mathematical disciplines in the first-year curriculum (curriculum 09.03.02.31).

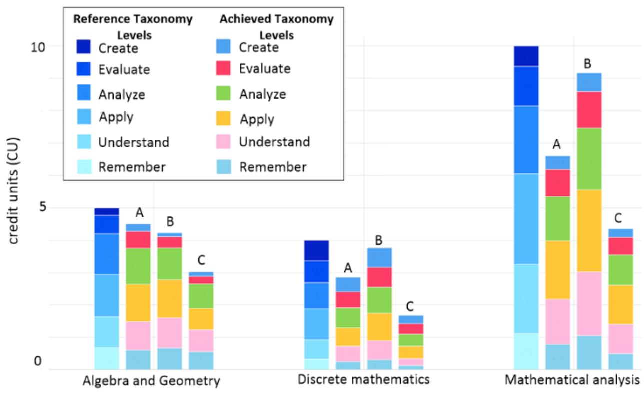

Fig. 9. Results of students A, B, and C, categorized by Bloom's Taxonomy levels, compared to planned outcomes (first-year mathematical disciplines of the curriculum)

A taxonomic analysis of student learning outcomes across courses allows for various conclusions regarding the formation of competency GPC-1. For instance, in the course *Algebra and Geometry*, Student C achieved relatively low scores. While his **Remember** level was quite acceptable, his **Understand** and **Apply** levels, as well as subsequent ones, differ significantly from the results of Students A and B. This indicates that he managed to memorize the basic concepts and mathematical structures of the modules on linear algebra, vector algebra, and analytical geometry, but encountered difficulties thereafter. This pattern is even more pronounced in Student C's results for other mathematical disciplines. Here, he likely found it much more challenging to grasp the fundamental concepts and methods of working with them, which most likely points to a weak mathematical background for this learner.

In contrast, Students A and B demonstrate a sufficiently high mastery of the disciplines. Student B only falls behind Student A in *Algebra and Geometry*, which may indirectly indicate the need to review the course assignments in more detail, potentially by introducing project-based tasks that would help students integrate various mathematical concepts into their future professional activities.

3. Discussion

A future specialist's professional qualification is developed through the progressive mastery of universal, general professional, and professional competencies, which corresponds to the achievement of the educational outcomes stipulated by the educational program's curriculum. Given the capabilities of the digital educational environment, including the continuous collection and storage of large volumes of educational data, it becomes possible to ensure complete transparency of the learning process. This allows for tracking the dynamics and decomposing all components of the educational results.

The transformation of educational outcomes can be represented (constructed and visualized) as a *competency ladder*. By ascending this ladder, the future specialist becomes a highly qualified professional, ready for modern occupational challenges. The proposed approach to analyzing the transformation process of educational outcomes (competencies) enables a detailed examination of the route (learning trajectory) a student follows, as well as a clear visualization of their intermediate results within the context of Bloom's Taxonomy levels. This facilitates the timely identification of difficulties a student encounters during their studies and helps to determine their nature. Consequently, this provides the student, educators, and potential employers with a more comprehensive understanding of the future graduate's professional qualities.

The developed methodology can be scaled for use in various educational institutions. On the one hand, the software systems from which data is exported for the study are among the most common in educational establishments. On the other hand, the initial data will have a similar structure and set of attributes, even if different software is used for managing educational programs and implementing the digital learning environment.

The proposed comprehensive approach for assessing educational outcomes imposes moderate computational requirements, typical for educational data mining tasks. The key operations include:

- Matrix factorization (PMF) for imputing missing grades, which requires resources proportional to the matrix size (number of students by number of assessment items) and has a computational complexity of $O(\mathrm{N\_Iter} \times (M + N) \times k^2)$, where k is the number of latent parameters (features) and $\mathrm{N\_Iter}$ is the number of algorithm iterations.

- Classification of tasks according to Bloom's Taxonomy levels, which is a deterministic matching operation against a reference matrix and does not require significant processing power.
- Data aggregation and normalization for calculating percentage shares and specific weights, which involves a series of arithmetic operations with matrices/vectors.
- Visualization of graph models and diagrams (treemaps, bar charts, etc.), typically performed using modern libraries; this work utilized Gephi and the BI system Yandex DataLens.

Thus, the approach does not require high-performance computing and can be deployed on standard server hardware or in a cloud environment with moderate resources. The main practical challenge lies not in computational power but in the correct integration of the developed software modules and ensuring data integrity during the consolidation phase. The primary processing load falls on the matrix factorization stage, which, however, is executed periodically rather than in real-time, ensuring the approach's scalability for the typical size of student groups and educational programs.

As previously noted, preparing highly qualified specialists is a complex task involving numerous factors. Therefore, the proposed approach to analyzing the transformation of students' educational outcomes can serve as a valuable tool for optimizing the educational process and managing its quality. It can also be utilized to develop an automated decision support system for the educational process, based on the analysis of learning data.

4. Conclusions

The presented technology for analyzing the transformation of student competencies in a university enables the decomposition of educational outcomes according to Bloom's taxonomy levels at various stages of the educational process. These stages encompass both temporal frameworks (semester, year of study) and the content of various disciplines and their associated modules. This comprehensive approach enhances the transparency of the educational process for all stakeholders, allowing for the clear structuring and assessment of educational outcomes, as well as the definition of the cognitive process levels students are expected to demonstrate.

The application of Bloom's Taxonomy facilitates the formulation of quality criteria for mastering individual disciplines and educational programs as a whole. Implementing the proposed approach will enable the timely identification of problems students encounter during their progression through the educational program. This will contribute to the creation of a flexible (adaptive) methodological system for subject-specific teaching. It is crucial for the educational institution to be able to respond promptly and create the necessary conditions for fostering professional competencies when the set goals are not being adequately achieved.

The implementation of this approach will be significant not only for managing the quality of future specialists' professional training but will also establish a foundation for developing a new methodology for the hybrid management of educational program quality. This will make it possible to build a unified analytical learning management system based on the operational analysis of educational data and will ensure the transformation of the educational process management model within the university.

References

1. Uvarov, A. Y, E. Gable, I. V. Dvoretskaya, I. M. Zaslavsky, I. A. Karlov, T. A. Mertsalova, P. A. Sergomanov, I. D. Frumin. Difficulties and Prospects of the Digital Transformation of Education. Moscow, PH HSE, 2019. 344 p. DOI:10.17323/978-5-7598-1990-5.
2. Romero, C., S. Ventura. Educational Data Mining and Learning Analytics: An Updated Survey. – WIREs Data Mining and Knowledge Discovery, Vol. **10**, 2020, p. e1355. DOI: 10.1002/widm.1355.
3. Stoyanov, S. N., T. A. Glushkova, A. G. Stoyanova-Doycheva, I. K. Krasteva. The Virtual Education Space: Concept, Architecture, Application. – Informatics and Education, Vol. **36**, 2021, No 9, pp. 47-54. DOI: 10.32517/0234-0453-2021-36-9-47-54.
4. Baig, M. I., L. Shuib, E. Yadegaridehkordi. Big Data in Education: A State of the Art, Limitations, and Future Research Directions. – International Journal of Educational Technology in Higher Education, Vol. **44**, 2020, No 17. DOI: 10.1186/s41239-020-00223-0.
5. Klašnja-Milićević, A., M. Ivanović, B. Vesin, M. Satratzemi, B. Wasson. Editorial: Learning Analytics – Trends and Challenges. – Frontiers in Artificial Intelligence, Vol. **5**, 2022, 856807. DOI: 10.3389/frai.2022.856807.
6. Semenov, A. L., A. E. Abylkassymova, T. A. Rudchenko. AI Methods in Control of Personalized General Education. – Doklady Mathematics, Vol. **109**, 2024, No 3, pp. 191-196. DOI: 10.1134/s1064562424702119.
7. Teasley, S. D., M. Kay, S. Elkins, J. Hammond. User-Centered Design for a Student-Facing Dashboard Grounded in Learning Theory. – In: Visualizations and Dashboards for Learning Analytics. Cham, Switzerland, Springer, 2021, pp. 191-212. DOI: 10.1007/978-3-030-81222-5.
8. Permana, A. A. J., G. A. Pradnyana. Recommendation Systems for Internship Places Using Artificial Intelligence Based on Competence. – Journal of Physics: Conference Series, Vol. **1165**, 2019. DOI: 10.1088/1742-6596/1165/1/012007.
9. Kustitskaya, T. A., R. V. Esin, Y. V. Vainshtein, M. V. Noskov. Hybrid Approach to Predicting Learning Success Based on Digital Educational History for Timely Identification of At-Risk Students. – Education Sciences, Vol. **14**, 2024, No 6, 675. DOI: 10.3390/educsci14060657.
10. Bloom, B. S., M. D. Engelhart, E. J. Furst, W. H. Hill, D. R. Krathwohl. Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain. New York, David McKay Company, 1956. 207 p.
11. De la Torre, J. DINA Model and Parameter Estimation: A Didactic. – Journal of Educational and Behavioral Statistics, Vol. **34**, 2009, No 1, pp. 115-130. DOI: 10.3102/1076998607309474.
12. Salakhutdinov, R., A. Mnih. Bayesian Probabilistic Matrix Factorization Using Markov Chain Monte Carlo. – In: Proc. of 25th International Conference on Machine Learning (ICLM'25), 2008, pp. 880-887. DOI: 10.1145/1390156.1390267.
13. Fusi, N., R. Sheth, H. M. Elibol. Probabilistic Matrix Factorization for Automated Machine Learning. – arXiv Preprint arXiv: 10.48550/arXiv.1705.05355, 2018.
14. Zhang, T., H. Mao, H. Liu, Y. Liu, M. Yu, W. Wu, G. Yu, B. Wei, Y. Guan. Parallel Prediction Method of Knowledge Proficiency Based on Bloom's Cognitive Theory. – Mathematics, Vol. **11**, 2023, 5002. DOI: 10.3390/math11245002.
15. Blagoev, I., G. Vassileva, V. Monov. A Model for e-Learning Based on the Knowledge of Learners. – Cybernetics and Information Technologies, Vol. **21**, 2021, No 2, pp. 121-135.

16. S e r d i t o v a, N. E., A. V. B e l o t s e r k o v s k y. Education, Quality, and the Digital Transformation. – Higher Education in Russia, Vol. **29**, 2020, No 4, pp. 9-15. DOI: 10.31992/0869-3617-2020-29-4-9-15.
17. K r o k i n s k a y a, O. K., S. Yu. T r a p i t s i n. Student as "An Education Consumer": Content of the Concept. – Higher Education in Russia, Vol. **6**, 2015, pp. 65-75.
18. Z y k o v a, T. V., A. A. K y t m a n o v, E. A. K h a l t u r i n, Y. V. V a y n s h t e y n, M. V. N o s k o v. The Algorithm for the Analysis and Evaluation of Educational Programs' Curricula. – Informatics and Education, Vol. **39**, 2024, No 1, pp. 52-64. DOI: 10.32517/0234-0453-2024-39-1-52-64.
19. Z y k o v a, T. V., A. A. K y t m a n o v, M. V. N o s k o v, E. A. K h a l t u r i n. Application of a Force-Directed Graph Drawing Algorithm for the Analysis of Curricula of Educational Programs of Higher Education. – Modern Information Technologies and IT-Education, Vol. **19**, 2023, No 1, pp. 104-116. DOI: 10.25559/SITITO.019.202301.104-116.
20. Z y k o v a, T. V. Applying Bloom's Taxonomy for Classifying Learning Outcomes in an Electronic Information and Educational Environment. – Open Education, Vol. **29**, 2025, No 4, pp. 55-63. DOI: 10.21686/1818-4243-2025-4-55-63.
21. A l m á a i t a h, W. Z., A. Q u r a a n, F. N. A L-A s w a d i, R. S. A l k h a w a l d e h, M. A l a z a b, A. A w a j a n. Integration Approaches for Heterogeneous Big Data: A Survey. – Cybernetics and Information Technologies, Vol. **24**, 2024, No 1, pp. 3-20.
22. K u s t i t s k a y a, T. A., R. V. E s i n, A. A. K y t m a n o v, T. V. Z y k o v a. Designing an Educational Database in a Higher Educational Institution for the Data-Driven Management of the Educational Process. – Education Sciences, Vol. **13**, 2023, No 9, 947. DOI: 10.3390/educsci13090947.
23. P a t e r e k, A. Improving Regularized Singular Value Decomposition for Collaborative Filtering. – In: Proc. of KDD Cup and Workshop, 2007, pp. 5-12.
24. M n i h, A., R. R. S a l a k h u t d i n o v. Probabilistic Matrix Factorization. – Adv. Neural Inf. Process. Syst., Vol. **20**, 2007, pp. 1257-1264.
25. M o o n, T. K. The Expectation-Maximization Algorithm. – IEEE Signal Process. Mag., Vol. **6**, 1996, pp. 47-60.
26. C a r l i n, B. P., S. C h i b. Bayesian Model Choice via Markov Chain Monte Carlo Methods. – Journal of the Royal Statistical Society, Vol. **3**, 1995, pp. 473-484.
27. W h i t e, H. Maximum Likelihood Estimation of Misspecified Models. – Econometrica, Vol. **50**, 1982, No 1, pp. 1-25.

Received: 17.10.2025, First revision: 04.12.2025, Second revision: 08.12.2025, Accepted: 14.12.2025