Waldemar Wojcik[1], Saule Smailova[2], Indira Uvalieva[3]

DEVELOPMENT OF A COMPREHENSIVE EDUCATION DATA ANALYSIS TECHNOLOGY

The present article offers a comprehensive technology for education data analysis which combines monitoring methods with the methods of information and mathematical support for administrative decision making in higher professional education.

Keywords: *data analysis; decision-making methods; monitoring of customer satisfaction; OLAP-analysis; fuzzy sets theory.*

Вальдемар Вуйчик, Саулє Смаїлова, Індіра Увалієва

РОЗРОБКА КОМПЛЕКСНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ ДАНИХ ЩОДО ОСВІТИ

У статті запропоновано комплексну технологію аналізу даних щодо освіти, що об'єднує методи моніторингу, інформаційної та аналітичної підтримки прийняття управлінських рішень у сфері вищої професійної освіти.

Ключові слова: *аналіз даних; прийняття рішень; моніторинг задоволеності споживачів; OLAP-аналіз; теорія нечітких множин.*

Рис. 5. Літ. 15.

Вальдемар Вуйчик, Сауле Смаилова, Индира Увалиева

РАЗРАБОТКА КОМПЛЕКСНОЙ ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ ПО ОБРАЗОВАНИЮ

В статье предложена комплексная технология анализа данных по образованию, объединяющая методы мониторинга, информационной и математической поддержки принятия управленческих решений в сфере высшего профессионального образования.

Ключевые слова: *анализ данных; методы принятия решений; мониторинг удовлетворенности потребителей; OLAP-анализ; теория нечетких множеств.*

Problem statement. The performed analysis of scientific research and development in the field of creation of automated information systems for educational processes management (Vasiliev, 1990; Linkov and Kulagin, 2002; Naykhanova and Dambaeva, 2004; Sauanova, 2010) has shown that the majority of them are aimed at the development of methods and approaches to analysis and administration of education activities of an educational institution and at creation of automated education process management systems.

In the majority of higher educational institutions in-house software solutions are developed. Such solutions arise gradually as the need for automation of education and managerial tasks are formed. Corporate information systems rank first among software solutions, they are mostly oriented at automation of educational process management; software solutions for automation of business processes rank second.

The existing mathematical methods for analysis of information allow obtaining more interesting and relevant data than average indices, that may be used at all levels when managing an educational institution.

Data mining provides for the solution for the following tasks:

[1] Lublin University of Technology, Poland.
[2] Serikbaev East Kazakhstan State Technical University, Ust-Kamenogorsk, Kazakhstan.
[3] Serikbaev East Kazakhstan State Technical University, Ust-Kamenogorsk, Kazakhstan.

- disclosure of a large amount of knowledge known as guesswork, or not known at all before the use of data mining;
- detection of problematic situations at the level of data (e.g., poor academic performance) and their causes;
- operational analysis of large data volumes (e.g., results of ongoing monitoring during semester), and detection of deviations from the normal course;
- analysis of historical data in order to improve the educational processes in the future.

Intellectualization of information systems in the area of higher education is a promising development direction that can solve problems related to quality of education services.

At the same time, the researched issue is the development of methods and models for data analysis in education systems, as well as the deployment of an information subsystem that would implement the mentioned methods and models, which determines the relevance of the selected area of research.

Educational process management system is the object of this research. Methods and models of data analysis that improve the efficiency of educational process management of a higher education institution are of a particular interest.

Recent research and publications analysis. Currently, the methods of data mining have become widespread in various fields. Such scientists as A.A. Barsegjan and M.S. Kupriyanov (2004), G. Pyatetskii-Shapiro (2009), J. Han and M. Kamber (2001), H.C. Romesburg (2004) have carried out research in this area. Problems related to the educational processes data analysis have been considered in works by R. Baker and K. Yacef (2009), L.I. Grigoriev (2008) and some others.

Unresolved issues. A situation when large volumes of information that are practically not processed aggregate within information systems is a characteristic sign of modern management systems. The complexity of solving this problem consists in the necessity to use intelligent information processing algorithms, that could give visual and comprehensible results for decision making in order to improve education processes. Traditional information analysis methods do not allow solving the problems of this class.

Specificity of data analysis in the area of education lies in the fact that the majority of data analysis problems imply the development of models that reveal patterns within data. Therefore, descriptive mining models are the most popular one in the area of education. These tasks give consideration to improved understanding of the data being analysed. The key point in these models is the lightness and transparency of the results for human perception. Perhaps, the observed regularities will be a specific feature of particular data analysed and will not appear anywhere else, but it may still be useful, and therefore must be known. Clusterization and association rule mining are referred to as the tasks of such type.

At present, forecasting problem in the field of education has less weight than the problem of description. This could be explained by the fact that the education system itself changes quite rapidly and there is a significant number of indirect factors within it.

An important aspect in the choice of models is their transparency, which is especially important in education. Transparent models, the meaning of which can be understood without possessing any special knowledge, can be used to describe regularities.

Intellectualization of information system in education involves the development of specific education data analysis models that are based on classical statistical approaches (Soltan et al., 2013).

Results of the research performed suggest that a comprehensive approach is required for the educational process management based on monitoring of educational process, the use of operational techniques, data mining and mathematical methods of decision-making support.

Based on the results of the education management problems analysis, the purpose of the research has been identified: improving the efficiency of educational processes management based on the development of a comprehensive technology which combines monitoring methods with the methods of information and mathematical support for administrative decision making in the area of higher professional education. The proposed concept includes monitoring of customer satisfaction, monitoring of educational processes, methods of operational analysis and data mining, as well as relative assessment of academic performance based on the multicriteria choice of alternatives in a fuzzy preference relation.

Key research findings.

1. Development of the concept for comprehensive education data analysis technology. In order to ensure efficient operations of a quality management system and of an educational institution in general, information environment needs to be created that would allow managing processes, data and people. A higher educational institution's information environment is a projection of business processes of this institution onto the area of information technologies. There are 3 ways to create information systems for a higher educational institution: development based on ERP-systems, acquisition of ready-made developments and in-house developments.

With a view to increase the efficiency of educational processes management, the present research offers the following approaches:

- in terms of implementing ISO standards and TQM methodology (quality is defined by customers);
- based on information technology, which gives new data analysis capabilities to support decision making;
- application of decision-making methods based on the fuzzy sets theory.

Monitoring of customer satisfaction with education based on the methods of questioning is used in the research for the purpose of implementation of the first approach. In accordance with ISO 9001-2008, an organization shall determine and implement effective arrangements to ensure communication with its customers. A developed, documented and implemented procedure for collection and evaluation of customer satisfaction information may be a part of such arrangements.

Implementation of the educational process management approach based on information technologies ensures more efficient use of university resources to perform monitoring of educational process using a unified quality criteria system, to carry out prompt response to changes in external and internal environments, and to improve organisational structure of university administration with the purpose to ensure its interaction at various levels, determining strategic goals and objectives of the organization in the area of quality, also evaluating the satisfaction of internal and external customers etc. (Soifer et al, 2006). Within the framework of IT application

operational analysis of data and data mining were implemented in the present research, allowing the detection of hidden regularities in educational process data.

Peculiarity of the third approach stems from the fact that educational processes management is classified as a hard-to-formalize and a multifactor problem. Using the apparatus of the fuzzy sets theory to assess the quality of educational process would allow considering avoidable uncertainties, measuring them quantitatively, and, consequently, obtaining a more accurate evaluation.

Thus, the relevance of this research is determined by insufficient development of methods for targeted and continuous improvement of characteristics of the educational system that would take into account the multicriteriality of tasks being addressed and the inherent uncertainty of data.

Therefore, the comprehensive technology of education data analysis is a technology comprising operational analysis and data mining related to educational process data, methods of monitoring customer satisfaction and educational processes monitoring, as well as the methods of decision making based on the fuzzy sets theory.

Implementation of a comprehensive technology involves the processes of collecting and processing education process data. Data-flow diagram of a comprehensive education data analysis technology is shown in Figure 1.

The diagram shows the interaction of processes in monitoring the customer satisfaction, educational processes status and the data analysis process.

2. Operational analysis and data mining for the educational processes data. Information systems of higher educational institutions in Kazakhstan accumulate large amounts of information on education activities of students. This data is mostly used for statistical reporting purposes and when ranking students according to their total average score obtained during the entire education period.

It proves necessary to use operational data analysis and intelligent information processing algorithms, that could give visual and comprehensible results for decision making to improve the education processes. However, certain problems remain unsolved, such as the issues of application of data mining methods for data analysis and decision making in the area of education (Baker and Yacef, 2013).

OLAP-analysis results allow for timely and reasoned decision making in educational processes management.

This paper presents a methodology for data mining implementation on educational processes data in order to search for large amounts of data related to non-obvious, objective and practically useful regularities. The scheme of the proposed methodology is shown in Figure 2.

3. Development of an algorithm for analysis of academic performance data based on the multicriteria choice of alternatives. Following the implementation of the rating system in higher educational institutions, the amount of information on educational processes status has increased significantly. The task of rating data analysis and processing becomes essential, as does interpretation of the results obtained to ensure management of the above processes.

In order to solve the tasks of streamlining data under the conditions of fuzzy initial information, fuzzy data analysis methods were used, as proposed by L.A. Zadeh (1976).

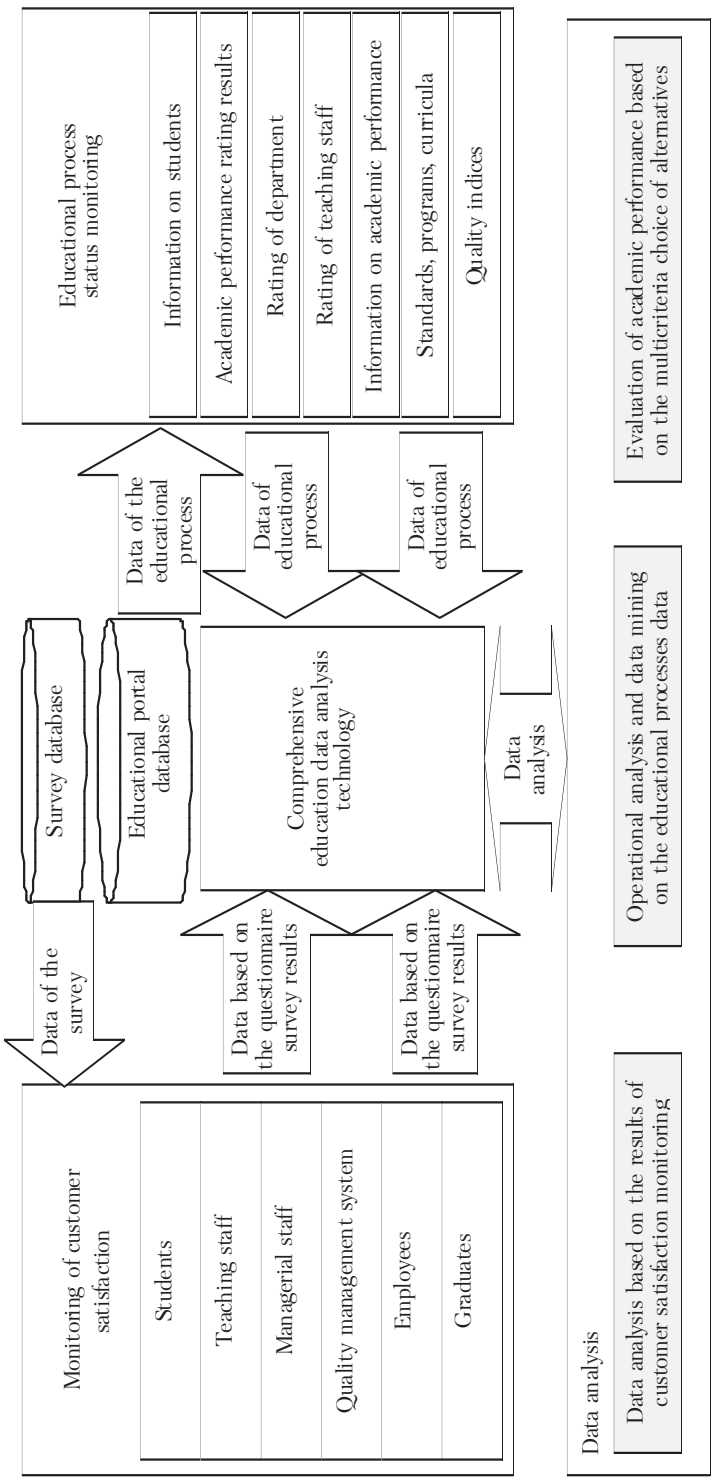
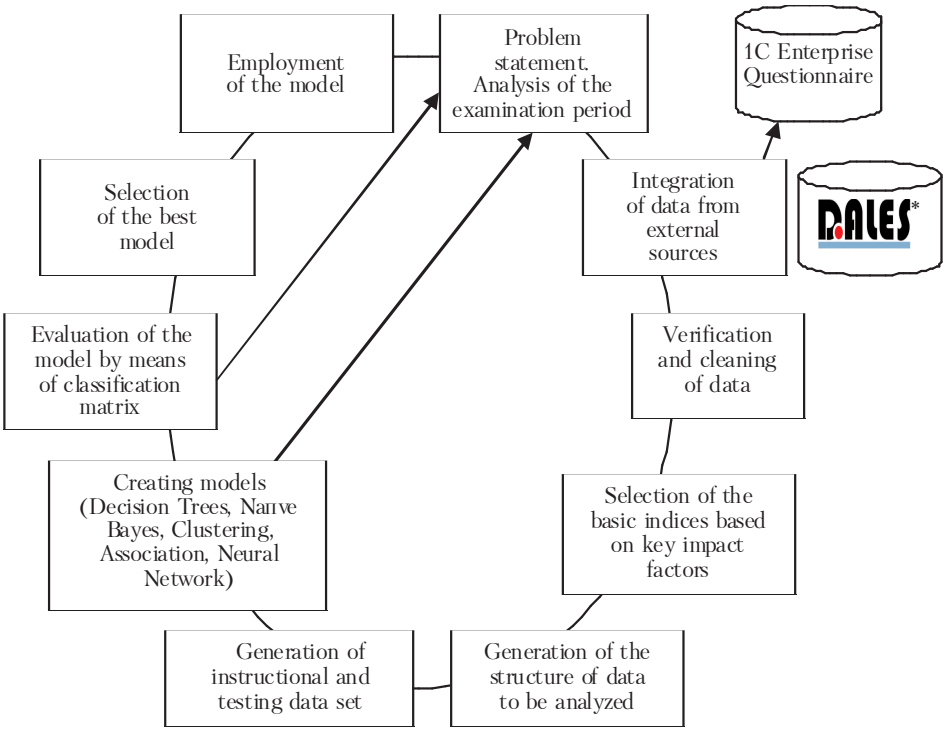


Figure 1. **Dataflow diagram of the comprehensive education data analysis technology,** *authors' development*



* Dales – Serikbayev East Kazakhstan State Technical University educational portal database.

Figure 2. **Methodology of data mining implementation on education data,** *authors' development*

S.A. Orlovsky (1981) suggested the techniques developed to process fuzzy information given in the form of fuzzy sets, which lead to new formulations of optimization and game problems. A.N. Borisov et al. (1990) offered a decision-making method based on the development of a set of non-dominated alternatives based on fuzzy preference relations.

These methods make it possible to solve the task of relative assessment of students' academic performance in various groups and to rank students of one group based on the sets of subjects.

Algorithm scheme for evaluation of academic performance based on the multicriteria choice of alternatives is shown in Figure 3.

According to the given algorithm, the following serve as the input data: a set of alternatives $\{x_i\}$, set of attributes of each alternative $\{a_i\}$, coefficients of relative importance of attributes $\{\lambda_k\}$.

Preference matrices Rk for each attribute are formed. Information on the pairwise comparison of alternatives for each feature is presented in the form of fuzzy preference relations (fuzzy relation "no worse" was introduced): $if x_i \geq x_j$, then $\mu_k(x_i x_j) = 1$, otherwise $\mu_k(x_i x_j) = 0$. After matrices are formed for each feature, a matrix is constructed for Q_1 fuzzy relation as intercrossing of all R_k.

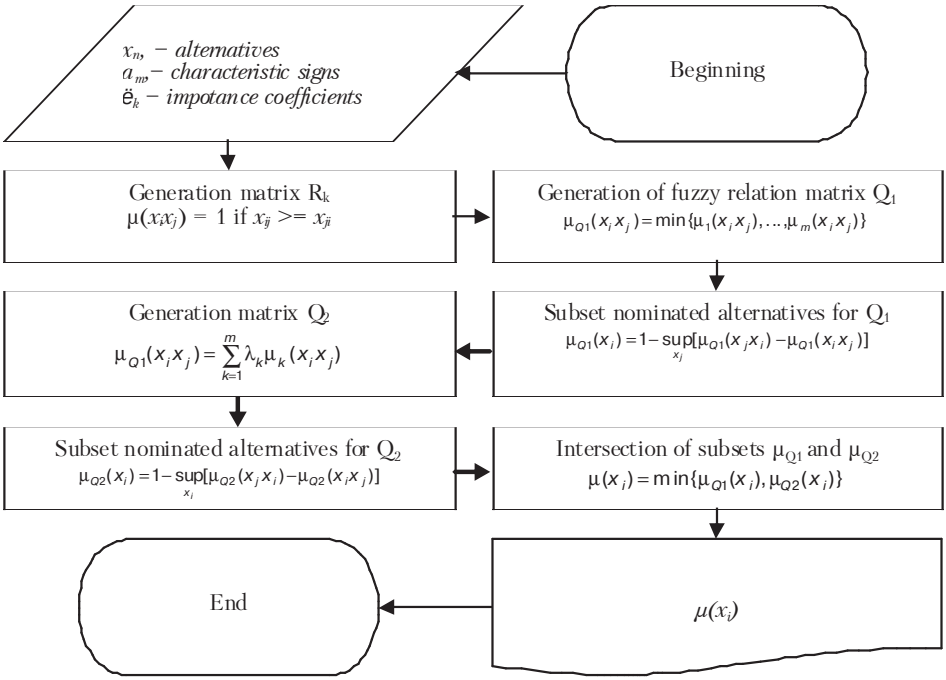


Figure 3. **Algorithm scheme for academic performance evaluation based on the multicriteria choice of alternatives,** *authors' development*

Then, a strict preference μ_{Q1}' fuzzy relation matrix is formed, and the degree of non-domination of alternatives is found according to Q_1 relation).

The Q_2 matrix is formed; a set of non-dominated alternatives is constructed for Q_2.

Intercrossing of μ_{Q1} and μ_{Q2} is the last step; it characterises the degree of an alternative's non-domination.

4. Software implementation of the comprehensive education data analysis technology. The proposed comprehensive technology for education data analysis was implemented in the form of an information analytical system (IAS). The purpose of this IAS is to increase the efficiency of education processes management in a higher educational institution through the use of monitoring facilities and intelligent methods to ensure decision-making support.

The outcomes of implementation of the comprehensive education data analysis technology prove useful not only for management-related and organisational decision making by managerial staff, but for students as well. The results will contain information that enables students objectively evaluate their own abilities and develop their own learning strategies. Carrying out operational analysis and data mining provides an opportunity to close the cycle of education process management.

Implementation of the information analytical system for educational process management (Figure 4) has a modular structure and contains the following 3 modules:

- Monitoring of customer satisfaction module – implements basic tasks related to the analysis of opinions of customers in the educational process;
- Operational analysis and data mining module – allows solving a number of tasks related to educational process;
- Multicriteria evaluation of academic performance module – carries out relative assessment of academic performance and ranking of students by various criteria.

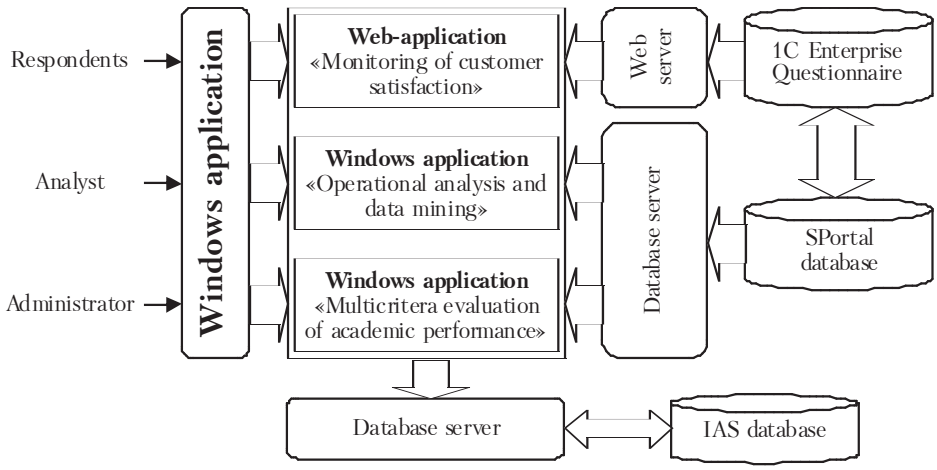


Figure 4. **Architecture of the IAS implementation,** *authors' development*

5. Key findings of the experimental study. 168 066 records have been created in the Operational Analysis and Data Mining Module to ensure experimental processing of data. Of those, "Instructional set" amounted to 100839, "Testing set" amounted to 40336, and "Forecasting data" – to 26891 records.

Data to be analysed were collected in IAD OP DB from the Serikbayev East Kazakhstan State Technical University educational portal database for 2009–2012. To ensure experimental processing of data, the following fields were considered for the analysis: department, department of graduation, year of admission, group, full name of student, specialty, year of study, education basis, year of study when the subject was studied, subject cycle, number of credits, language of study, UIT score, RK1, RK2, examination, school, type of high-school graduation certificate, high-school graduation certificate score, absence, professor, academic degree, academic rank, position, full-time/part-time, teaching and guiding, research work, educational work, research and development of the department, material and technical facilities of the department, involvement of foreign scientists.

Prior to the development of a model for data mining, a user is given an opportunity to perform the analysis of the key impact factors. The analysis of the key factors impacting the successful passing of the examination period has shown that the following factors have a more significant impact: absence, year of study, education basis, speciality, educational work.

An analyst has an opportunity to develop IAS models based on the structure built at the preceding stage and based on the following algorithms: decision-making tree, neural network, simplified Bayesian algorithm, logistical regression, time series.

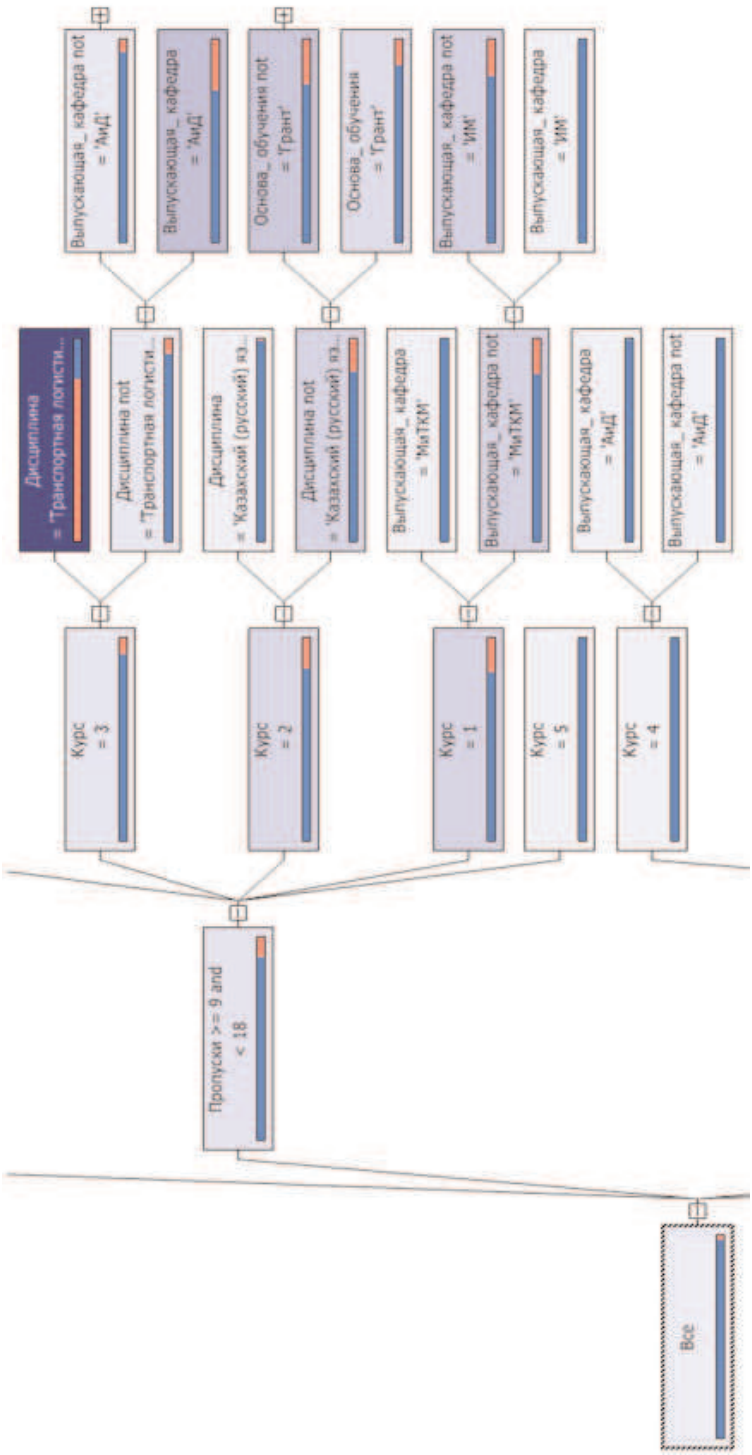


Figure 5. **Window of the Information and Analytical System including a decision tree fragment,** *authors' development*

Educational process data modelling results are presented above. Figure 5 shows the results of classification based on a decision-tree algorithm, where probability for an option (passed/not passed) can be seen for each node. A developed decision-making tree and a dependency diagram are shown for the created model. The decision-making tree shown in Figure 6 allows assessing the developed model. There is also a possibility to set the number of tree levels displayed.

Conclusions. Therefore, the following major scientific and practical research results have been obtained:

- a concept for comprehensive education data analysis technology has been developed; it includes monitoring facilities in the area of education and intelligent methods of decision-making support;
- a methodology for educational process data analysis has been developed; it is based on methods and models of operational analysis and data mining;
- an algorithm for relevant evaluation has been proposed to evaluate academic performance based on the multicriteria choice of alternatives;
- architecture of information analytical system of educational process management has been proposed.

References:

Борисов А.Н., Крумберг О.А., Федоров И.П. Принятие решений на основе нечетких моделей. – Рига: Зинатне, 1990. – 187 с.

Васильев В.Н. Модели управления вузом на основе информационных технологий: Автореф. дис... докт. техн. наук. – Петрозаводск. 2001. – 108 с.

Григорьев Л.И. Научно-методические и технологические основы информационной системы управления качеством учебного процесса. – М.: Нефть и газ, РГУ нефти и газа им. И.М. Губкина, 2008. – 132 с.

Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений / Пер. с англ. – М.: Мир, 1976. – 165 с.

Линьков В.М., Кулагин В.П. Проблемы автоматизации задач управления учебным процессом в автоматизированной информационной системе вуза // Индустрия образования: Сборник статей. – Вып. 2. – М.: МГИУ, 2002. – С. 162–170.

Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов и др. – СПб.: БХВ-Петербург, 2004. – 336 с.

Найханова Л.В., Дамбаева С.В. Методы и алгоритмы принятия решений в управлении учебным процессом в условиях неопределенности: Монография. – Улан-Удэ: ВСГТУ, 2004. – 164 с.

Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. – М.: Наука, 1981. – 155 с.

Пятецкий-Шапиро Г. Data Mining и перегрузка информацией: Вступительная статья к книге // Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. – 3-е изд. перераб. и доп. – СПб.: БХВ-Петербург, 2009. – С. 13–14.

Сауанова К.Т. Исследование и разработка моделей и методов системы обеспечения качества информационной образовательной среды: Автореф. дис... канд. техн. наук. – Алматы, 2010. – 23 с.

Сойфер В.А., Гречников Ф.В., Кузьмичев В.С., Колпаков В.А. и др. Система управления качеством образования в университете на основе информационных технологий // Управление качеством.– 2006.– №3. – С. 92–96.

Солтан Г., Смаилова С., Увалиева И. Информационно-аналитическая система управления качеством учебного процесса // Вычислительные и информационные технологии в науке, технике и образовании – 2013: Матер. Междунар. конф. – Усть-Каменогорск, 2013. – С. 326–329.

Baker, R.S.J.d., Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1(1).

Han, J., Kamber, M. (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers.

Romesburg, H.C. (2004). Cluster Analysis for Researchers. Krieger Pub. International Conference on Computer Science and Software Engineering, Washington, DC. Pp. 452–455.

Стаття надійшла до редакції 17.06.2014.