UDC 004.896

Intelligent Data Analysis in Quality Management Problems of Education Process

Eastern-Kazakhstan State Technic University n.a. D. Serikbaev
G.J. Soltan, S.S. Smailova, I.M Uvalieva
National Research Tomsk Polytechnic University
A.K. Tomilin

G.J. Soltan

S.S. Smailova

I.M. Uvalieva

A.K. Tomilin

The article describes the Intelligent Data Analysis (IDA) system model applied to the University education process, the possiblities of IDA in consideration of the characteristic education aspects and its application.

Key words: *educational process, quality management system, data analysis.*

For effective operation of university quality management system it is necessary to establish an IT environment, which, in its turn, could control the process of capturing and data analysis. University IT environment is the representation of university process intelligence within the framework of IT. According to [1], quality management system within any university cannot be considered to be integral and effective without applying those tools which would monitor the education process itself through IT.

Thus, the research target is to apply Intelligent Data Analysis IDA methods in the quality management problem-solving of the education process.

One of the eight principles of the International Standard ISO 9000[2] is the following approach to decision-making based only on facts. The implementation of this principle involves the analysis and capturing of reliable and accurate data relevant to the assigned task.

Capturing and further data analysis involves a definite possession of knowledge and use of specific methods. One of these methods is the so-called Intelligent Data Analysis (IDA), providing more interesting data rather than in the case of obtained average data [3].

Intelligent Data Analysis (IDA) of the education process allows solving the following problems:

- defining the cluster of students being in the risk group of academic progress;
- on-the-spot analyzing enormous volume of data (for example, the results of current and final assessment) and detecting departure from normally running sequence;
- detecting problem situations in the education process and defining their causes;
- analyzing accumulated data to improve the education process.

Characteristic aspects of data analysis within the education system involves the fact that solving such problems includes modeling, which in its turn, reveals the data pattern itself. In this case, the descriptive model has become one of the most in-demand models which facilitate a deeper understanding of the analyzed data. The key point of these model results is that they are simple and transparent for

human perception. It is probable that the defined patterns are characteristic for specific experimental data, regardless of the fact that this data may not be encountered anywhere, but may be practical and so must be studied. Clustering and search of association rules are classified as such kind of tasks [4].

At the present moment the forecasting problem in the education system is of less significance than that of a problem description. This is closely related to two facts: the education system itself is changing quite rapidly and there exist a great number of indirect factors.

Effective implementation of Intelligent Data Analysis (IDA) exists through free software products as back code implementation from common algorithm or stand-alone (SAS Enterprise Miner, Poly Analyst, Deductor, RapidMiner), as well as different tools integrated into data base management system – DBMS (Oracle Data Mining, SQL Server Analysis Services).

Applying off-the-shelf algorithms as back code is rather time-consuming, and in case of applying IT system stand alone applications, data selection and cleaning will be executed in on-line mode and data is exported in the format which is supportable for the external toolkit. Further data manipulating, visualizing and applying analysis method occur independently from the IT system. In this case, it is considered that the most appropriate method is those tools integrated into DBMS. The advantages of this approach are the following:

- possible integration of multivendor data;
- application of built-in data origination facility to intelligent analysis;
- selection and application of different IDA algorithms;
- store and reusability of models;
- possible comparative analysis of different effective models.

To implement the quality management tasks of the education process, an analytic education portal subsystem was designed at Eastern-Kazakhstan State Technic University (EKSTU), n.a. D. Serikbaev, which in its turn, provided the possible analysis through the tool Data Mining [6]. This subsystem included the following process and functions:

- compilation of education process statistics;
- designing multidimensional IDA database;
- data consolidation from different independently developed codes to analytic subsystem database and its management and application to perform online information-analytic and intelligent analysis;
- IDA implementation;
- preparation and presentation of IDA results for the analyst (or management representative);
- administrator access rights to information-analytic subsystem resources.

During the period of 2009-2013 168 066 write-records were generated for processing experimental data within the framework of the model "Online information-analytic and intelligent analysis", including 100 839 write-records in "Training set"; 40 336 write-records in "Testing set" and 26 891write-records in "Planning data".

The following IDA tools were applied in the analysis:

- decision-tree algorithm;
- neural net algorithm;
- simple Bayes algorithm;
- clustering algorithm;
- logistic regression algorithm [5].

IDA model defines what combinations of input variables trigger either high or low academic progress, which in its turn, isolates the student risk groups.

The results of education process data modeling is depicted below, i.e. Fig.1 shows the classification results of decision-tree algorithm.

Fig. 2 shows a window with dependency network for decision-tree solutions to define those factors influencing the results of semester-exams.

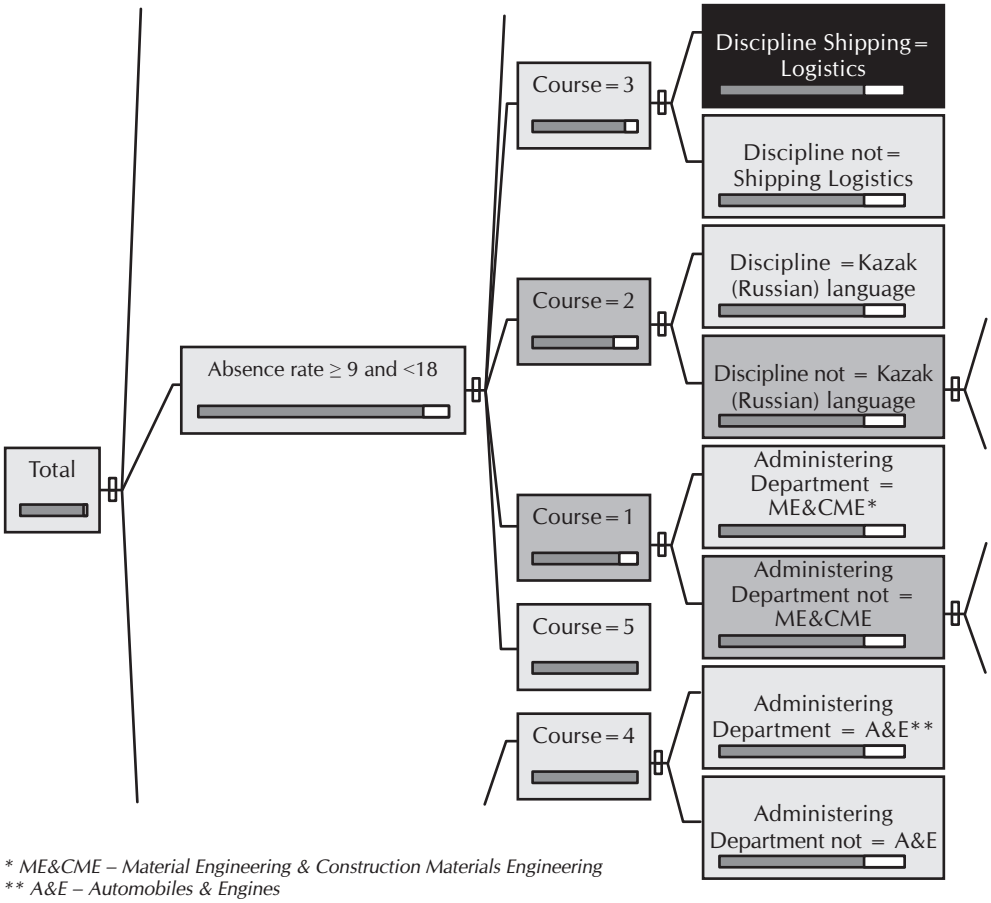
Based on the algorithm results of designed dependency network for decision-tree the following factors were defined which influence the results of semester-exams: discipline, administering department, absence rate, based-learning and course. The most significant factor was "Absence rate."

Fig. 3 shows the IDA results based on the neural net algorithms

The results of the classification based on Bayes algorithm is depicted in Fig. 4. According to the report of attribute comparison based on Bayes algorithm, the major factor influencing "failed semester exam" is the "absence rate."

According to Fig. 5, dependency network for Bayes algorithm showed that there are more factors influencing "semester-exam: passed" in comparison to the dependency network based on the decision-tree solutions. Based on the algorithm results of designed dependency network for Bayes model the following factors were defined, which influences the results of semester-exams: discipline, number of credits, absence rate, based-learning, course and learning language. Thus, the IDA of the education process was conducted based on the implementation of the following

Fig.1. Window with Fragment of Decision – Tree, Where Every Block Shows Probability of Variation (Passed \ Failed)



** ME&CME – Material Engineering & Construction Materials Engineering*
*** A&E – Automobiles & Engines*

ENGINEERING EDUCATION 13'2013

Fig. 2. Algorithm Results of Designed Dependency Network for Decision-Tree Solutions

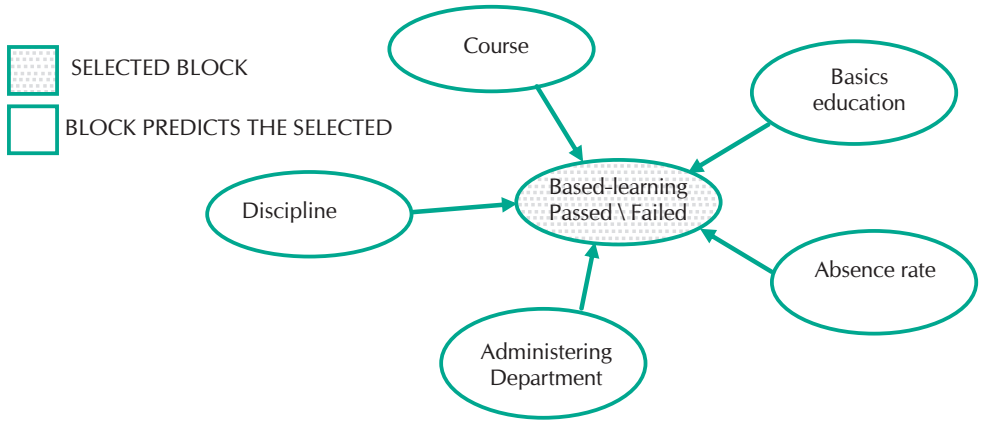


Fig. 3. Results of Classification through Neural Net Algorithm

Attribute	Notion	Failed
Discipline	History of Economic Science	
Speciality	50304	
Discipline	Money, Credits, Banks	
Discipline	Statistics	
Discipline	Shipping Logistics	
Discipline	Economic Data Processing	
Speciality	50724	
Speciality	5B090100	
Speciality	5B050700	
Discipline	Fundamentals of Electric Engineering	
Administering Department	Economic Law (EL)	
Speciality	5B050900	
Speciality	5B050600	
Speciality	5B050800	
Speciality	50806	
Faculty	Finance & International Trade (FIT)	
Speciality	50507	
Speciality	50508	
Discipline	Financial Markets and Brokers	
Discipline	Advanced Course in Mathematics	

algorithms: decision-tree, neural net and simple Bayes model.

The above-mentioned results of different analysis types could be practical not only for administrators in solving management and organization issues, but also for students, who would be able to reasonably predetermine his\ her abilities and develop his\her own learning strategies. Introducing on-line IDA facilitates the possible closed-loop management cycle of the education process itself.

Information-analytic subsystem architecture for any university is depicted in Fig. 6.

In conclusion, the information-analytic system of university education process quality management should not only provide information about the current and future processes, but also define those problem areas within the university education process itself, develop corrective actions to intensify the above-mentioned result-oriented processes. Corrective actions involve the results of online and intelligent processing, which in its turn, identifies the quality of the education process and propose hypotheses relevant to those methods improving this process.

Fig. 4. Results of Classification Based on Bayes Algorithm

Attributes	Notion	Passed	Failed
Absence rate	0		
Course	4		
Course	1		
Course	2		
Speciality	5B042000		
Based-learning	Contract		
Based-learning	Grant		
Speciality	5B090100		
Absence rate	20		
Absence rate	23		
Absence rate	27		
Absence rate	18		
Course	5		
Absence rate	25		
Absence rate	14		
Absence rate	12		
Speciality	50729		
Absence rate	17		
Absence rate	10		
Absence rate	15		

Fig. 5. Results of Plotting Dependence Grid for Bayes Algorithm

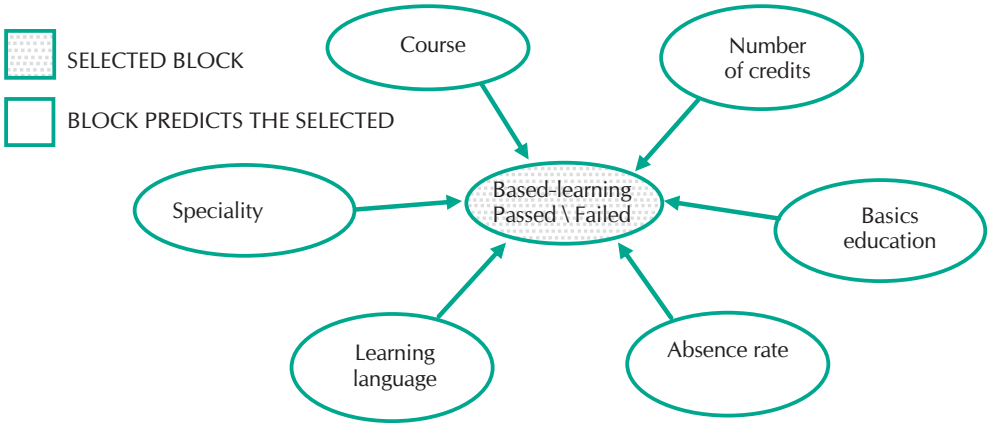
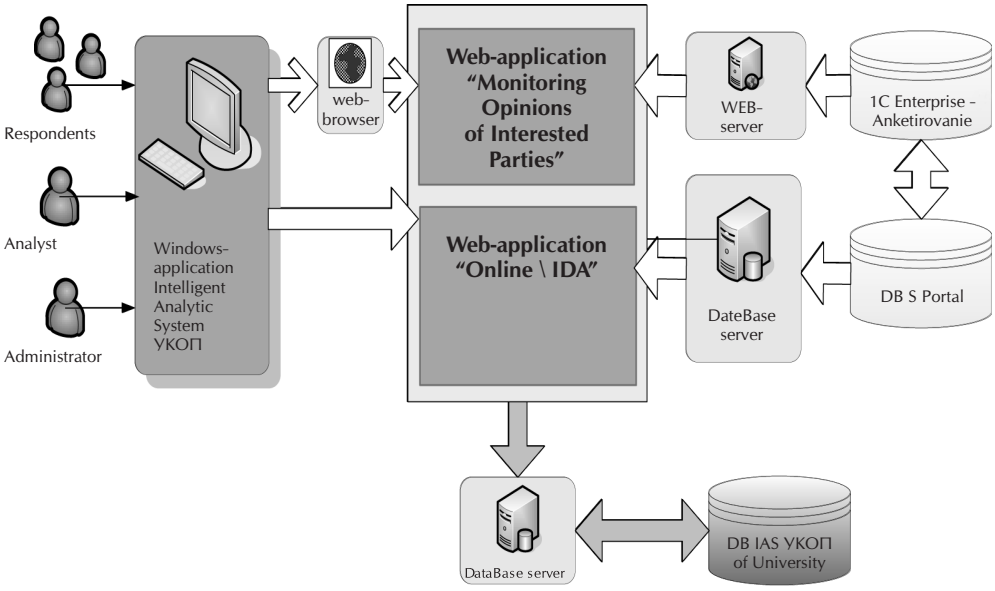


Fig. 6. Analytic Subsystem Architecture



REFERENCES:

1. Quality management in higher education institution. / Ed. Mutanov G., Tomilin A., Kukina Yu. (and others). – Ust-Kamenogorsk, 2011. – P. 116.
2. Standards of Kazakhstan Republic ISO 9000-2007. Quality Management System. Fundamental concepts and glossary. Introduction 2008–01–07. – Astana, 2007 – P. 76.
3. Konstantinovski D. From statistic data compilation to information application of decision-making. / Ed. Konstantinovski D., Agranovish M., Dimarskaya O.M. – M., 2006, – P. 206.
4. Methodological and technological background of IT for quality education process management. Ed. Grigoreva L. M. – M., 2008. – P. 132.
5. Methods and models of data analysis: OLAP и Data Mining Ed. Barsegyan A., Kupriyanov M., Stepanenko V., Kholod I. – St. Petersburgh, 2004. – P. 336
6. Tomilin A. (2012) In-home quality control of education in University. Enginnering Education. – 2012. – № 9. – pp. 56–61.