



IMPLEMENTATION OF DATA MINING ANALYSIS INTO THE LESSONS LEARNED SYSTEM IN PRIMARY EDUCATION

Julijana Vasiljević¹,
Dragan Vasiljević¹

¹University of Nis,
Faculty of Electronic Engineering,
Nis, Serbia

Abstract:

Changing people's function in the education system of the Republic of Serbia, as the first important step in the implementation of lessons learned (herein-after: LL), would allow educators in primary education system to admit and openly discuss the perceived mistakes made during the working process and also share what has been observed or done with other educational units within or outside the education system of the Republic of Serbia. It is an essential prerequisite for the efficiency and implementation of the LL. The achievement of the above involves the development of the capacity for self-analysis in an atmosphere where there is no danger of being condemned or criticized for making mistakes - mistakes should not be ignored if we want to avoid a decline in educational achievement.

This paper describes the basic functions and gives a suggestion for software functionality LL with the aim of improving the educational process and assisting in the professional orientation of students' basic education.

Also, the paper gives the results obtained by the author, analyzed by the software program „SPSS“ (Statistical Package for the Social Sciences) using the techniques of data mining as well as binary logistic regression.

Keywords:

basic education, data mining, binary logistic regression, information transfer.

1. INTRODUCTION

The school system of primary education in the Republic of Serbia is not sufficiently prepared to be changed and to achieve the goals set by the reformers. The main flaw is the inability of reformers to counteract this trend. A proof of such allegations are the results of „PISA“ (Programme for International Student Assessment) tests carried out in the Republic of Serbia, which assume constantly „poor results“, and therefore suggest that the quality of education is not at the expected level.

Results of the last PISA tests show that more than a third of 15-year-olds are functionally illiterate. Nearly 40 percent of students are below the level of mathematical functional literacy, a third do not meet the minimum functional linguistic knowledge, while 35 percent of scientific functional literacy. From all the above, it can be concluded that the problems in basic

Correspondence:

e-mail:
julija2921968@gmail.com



education in the Republic of Serbia have not been considered so that lessons are identified as a negative practice to be processed through the system whose product could be „lessons learned“.

2. DEVELOPMENT OF LL IN THE SYSTEM OF PRIMARY EDUCATION

The basis for further monitoring and understanding the importance of LL in the Primary Education System of the Republic of Serbia (hereinafter: PES RS) is the understanding of the concepts of „lesson“ and „lesson learned“.

Lesson knowledge or understanding is gained through experience. The experience can be positive (best practices) as a successful test, a mission, practice, or workshop, or negative, such as accidents or failures. Successes and even failures can be considered as a source of lessons.

The lesson is learned when there is a change in behavior. Obviously, this change in behavior has to be a positive one, which should improve the performance of the system considered.

There are some very serious reasons why PES RS needs to develop its capacity through effective LL. Some of the reasons are:

- ♦ It saves time, since it provides a central location for storing and enables efficient retrieval of valuable information.
- ♦ Reducing or avoiding the costs of providing information about the success or mistakes that can be avoided.
- ♦ The possibility of extending information network providing the ability to share information and connect with other sites, „experts“ or people doing similar things.
- ♦ Most importantly, it reduces the risk of repeating mistakes and increases the chance of success.

3. PROPOSAL STRUCTURE LL PES RS

Program LL can be a mechanism linking the existing range of tools that facilitate learning and information transfer. An extensive network of individuals who contribute to the process of sharing knowledge through the system have the potential to use information technology quickly overcoming time-critical points in the LL and to provide access to relevant information to all interested users. This can be especially important in cases where the information relates to events that may have very

negative consequences on the functioning of the educational institution.

Although the observation is the first building block in the process of identifying good and bad practices or events, the educational institution usually cares if the lessons are identified as problematic. Results of the analysis in the development of lessons, best practices and recommendations for corrective action should help solve problems. Figure 1 shows the structure based on LL.

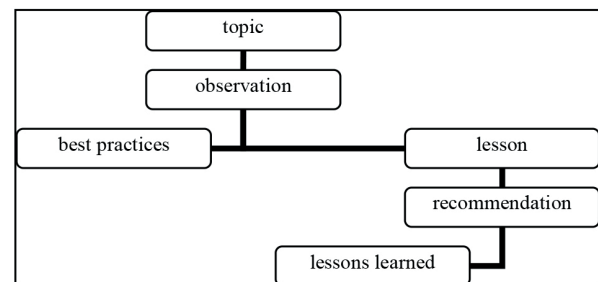


Fig. 1. Shows the basic structure of LL

Source: Author

A topic is a general theme that has been identified as a potential problem.

An observation is a statement based on experience or perceived during teaching.

A lesson represents a potential solution to the problem based on experience and observations of irregularities.

A recommendation represents corrective actions which, when taken, should solve the perceived problem.

The best practice is the solution to the problem, or actions, procedures and technical solutions that are used in work.

The lesson learned here is the statement that the implementation of corrective solutions leads to better results and results expected by the students.

Programs LL will vary according to the needs of the institution, but most will have at least six functions as follows:

- ♦ Data collection.
- ♦ Data analysis.
- ♦ Sharing of information.
- ♦ Archiving.
- ♦ Settlement and
- ♦ Assessment.

Algorithms of LL functionality are shown in Figure 2.

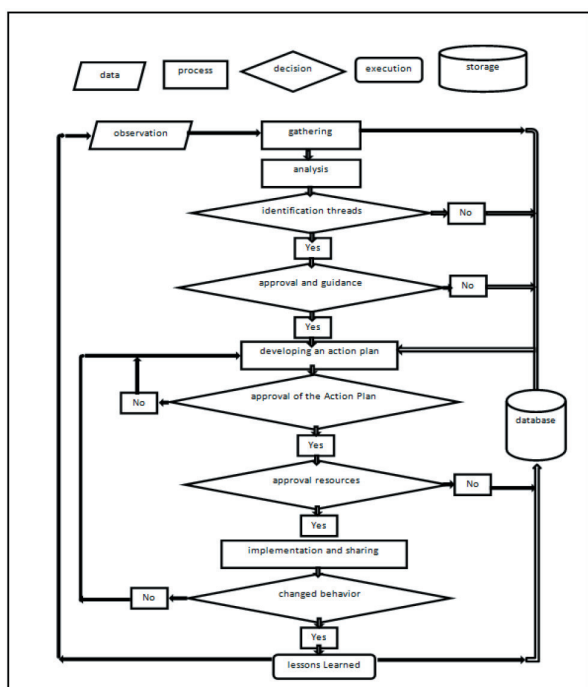


Fig. 2. Algorithms of LL functionality

Source: Author

Collecting data is the first function of a LL's ability to collect data relevant to the subject of the analysis. Options for the collection of data on issues in accordance with the institution's mission should be supported by the LL. Here is a list of potential options for information collection:

- ◆ Tasks and exercises made by students.
- ◆ Experiments and tests that support the teaching unit.
- ◆ Constant monitoring of the work and progress of students and
- ◆ Interviewing.

4. IMPLEMENTATION OF DATA MINING METHODS IN LESSONS LEARNED SYSTEM

Analysis is the process that is used for a thorough understanding of the field and has the function of recognizing the potential for improving the functioning of the institution.

For the purposes of research and implementation of data mining method of binary logistic regression, the collected data should be sorted into groups, as follows:

The first set of data is obtained from a survey of teaching staff of eight elementary schools from the territory of the City of Belgrade and the municipalities of Cukarica,

Obrenovac and Rakovica. Questionnaire - Likert scale paragraphs 1-5 (8 independent and 51 dependent variable). Total of 302 respondents (2 surveys are invalid), processed in the program tool „SPSS“ 300 surveys.

The second group of data is obtained from a survey of parents whose children attend primary school in the area of the City of Belgrade and the municipalities of Cukarica, Obrenovac and Rakovica. Questionnaire - Likert scale paragraphs 1-5 (7 independent and 32 dependent variables). A total of 320 respondents (20 poll will be void), processed in the program tool „SPSS“ 300 surveys.

The third group of data is the data:

- ◆ The total number of the employees of the teaching staff at eight elementary schools from the territory of the City of Belgrade and the municipalities of Cukarica, Obrenovac and Rakovica.
- ◆ The total number of pupils in these schools.
- ◆ The number of eighth-grade students at the end of the school year 2014/2015.
- ◆ The results achieved (grade point average) during training in mathematics and mother tongue and
- ◆ The results of the final exam (students' mastery of knowledge and skills).

After the data collection and transformation of data into the software package "SPSS" for the first and second sets of data, factor analysis was performed as follows:

Exploratory factor analysis and confirmatory factor analysis.

Exploratory factor analysis of the first group of data acquired through questionnaires answered by teaching staff of eight elementary schools set aside a total of 13 factors of 8 independent and 51 dependent variable, where the author of the paper, for the comparative analysis of the data of the first and second groups of data, lists 7 factors: FN1 - method of operation and organization of the scientific (MRON), FN2 - social factor (SF), FN3 - quality of work (KR), FN4 - the use of ICT in teaching (ICT), FN5 - syllabus (ON), FN6 – parents' engagement (AR) and FN7 - content textbooks (SU).

Confirmatory factor analysis was performed on data by extraction method „PCA“ - (Principal Component Analysis), which was verified by grouping variables, and which confirmed that these variables are essentially grouped around the above factors.

Variables which respondents gave with a similar attitude or sense of FN1 are: V3 - In your opinion, are the classes held successfully, V4 - In your opinion, your colleagues hold classes successfully, V8 - The success of the students from your course is at a high level, V10 - Physical



activity positively affects the learning ability of the students, V14 - Students are in classes, V15 - Students are aware of the application of lessons learned, V19 - Are you satisfied with the working conditions and V40 - Teaching by topics has advantages over teaching of particular cases.

The results of confirmatory factor analysis, total factor Explanation of variables: FN1 - the methods of teaching (MRON) is given in Table 1.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.759	46.992	46.992	3.759	46.992	46.992
2	0.913	11.409	58.401			
3	0.848	10.594	68.995			
4	0.771	9.638	78.633			
5	0.691	8.640	87.272			
6	0.589	7.363	94.635			
7	0.239	2.990	97.624			
8	0.190	2.376	100.000			

Table 1. Results of confirmatory factor analysis for teaching staff

Source: Author

The system method was factor analysis for the second group of data obtained in the survey of parents whose children attend primary school.

The results of confirmatory factor analysis are preserved by programming tool "SPSS" as a variable to build a model of binary logistic regression.

As an important factor for predicting the level of achievement the author of this paper carried out comparative analysis FN6 - involvement of parents (teachers) and FR6 - involvement of parents (parents' attitude) using binary logistic regression.

Confirmatory factor analysis confirms that the variables V16, V21, V22 grouped around a single factor, shown in Table 2.

Q16 - Do you have regular contact with the teaching staff of the school in which your child is being educated, V21 - You have proposed in any way that teachers improve the educational process and V22 - your child does not pay enough attention to teachers at school.

	Component
	1
V21	0.928
V22	0.982
V16	0.976

Table 2. Results of confirmatory factor analysis

Source: Author

5. OVERVIEW OF THE BASIC METHODS OF DATA MINING

Basic methods of data mining used by the algorithm analysis fall into two main groups (Han & Kamber Data 2001): DM method using supervised algorithms and unsupervised algorithms.

Supervised algorithms are those in which models are built using the data so as to know in advance the groups which the data belong to, and then on the basis of the constructed model predict the group to which the unknown data will belong. These gathering methods belong to the classification of data and regression methods.

Non-supervised algorithms are based on the given data forming groups of data, without prior knowledge about to which group data might belong. This set of methods belongs to the grouping and association rules.

Regression is a method that predicts the value of numeric attributes. Based on the known values of attributes of the given data we determine model parameters, and then on the basis of the model parameters we determine unknown attribute values of new data.

The regression model with its evaluation methods is suitable for areas where the classification codes are used to predict categorical (discrete, unsorted) variables. Regression is used to predict the missing or unavailable



numerical values before the tag class. Prediction refers to both types of predictions, numerical and class prediction.

Regression analysis is a statistical method that is used primarily for numerical forecasting, and as such includes the identification of trends over the available data (De Veaux, 2003; Goharian & Grossman, 2003).

There are several different types of regression in statistics, but the more general idea is to create a model that maps values of predictors in a way that the possibility of error in forecasting is the smallest. Basic types of regression are (Han & Kamber Pei, 2012; Ngai, et al., 2011): linear regression, multiple regression, logistic regression and nonlinear regression.

Linear regression is the simplest form of regression in which the value of the dependent random variable Y is defined as a linear function of the predictive variables X:

$$Y = \alpha X + \beta \quad (1)$$

where the regression coefficients α и β are determined using the method of least squares.

Given a set of samples is Y, wherein each pattern is shown as a two-dimensional vector (x_i, y_i) . Number of samples is y.

Ratios α и β are defined as:

$$\beta = \frac{\sum_{i=1}^u (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^u (x_i - \bar{x})^2}, \quad \alpha = \bar{y} - \beta \bar{x} \quad (2)$$

The values:

$$\bar{x} = \frac{\sum_{i=1}^u x_i}{u}, \quad \bar{y} = \frac{\sum_{i=1}^u y_i}{u} \quad (3)$$

Multiple regression is an extension of linear regression. In contrast to the linear regression analysis it involves more than one predictive variable. The dependent variable Y is defined as a linear function of the multidimensional vector of predictive variables. The general form of multiple regression is:

$$y = \alpha_1 x_1 + \dots \alpha_n x_n + \beta_1 \quad (4)$$

Where the parameters α_1, β_1 are determined by the method of least squares.

Logistic regression is an extension of linear regression with the restriction that the range of values that can take the dependent variable Y can only be in the range [0, 1].

The logistics function thus modeled as a linear function of the probability of predictive variables, the general form is:

$$y = \frac{1}{1 + e^{-x}} \quad (5)$$

If the variables in the models do not show linear than polynomial dependence, a polynomial dependence must be reduced to a linear form and the matter further as a linear regression.

Showing results of binary logistic regression on the attitude of parents to FR6 factor - the involvement of parents in relation to the level of achievement is shown in table 3.

Based on these results we can conclude that regular contact of parents with teaching staff at the school has a positive impact on improving the level of student achievement to a significant extent; the proposals on working methods will also have a positive impact on the improvement of the level, but to a lesser extent; the lack of attention from students the teachers greatly decreases the level of education (applies to gifted students and students with special needs).

Step	B	SE	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)		
							Low	Upper	
	V16	1.251	0.767	2.662	1	0.103	3.495	0.777	15.716
	V 21	0.148	0.285	0.271	1	0.603	1.160	0.664	2.026
	V 22	-1.387	0.771	3.233	1	0.072	0.250	0.055	1.133
	Cons	0.427	0.257	2.760	1	0.097	1.532		

Table 3. Showing results of binary logistic regression to the attitude of parents to factor FR6

Source: Author

Described above, and after the binary logistic regression indicates that the probability of the level of achievement for residents analyzed 0.41, and the same can be increased by increasing the activity of parents both at the individual and the general-school level.

6. FOR USER INTERFACE OF LL PES RS

The paper provided examples of „Open Source Technology“ for the implementation of the Proposals solutions LL PES RS, which can be used under Windows and Linux platforms, and their integration with LL PES RS will not affect the operation of LL PES RS by the proposed concept.

An important aspect of the solution each represents a system for managing databases such as MySQL, multi-threaded, multi-user SQL system. The system works as a server providing multi-user interface to access the database that is widely used as an open-source system for managing relational databases (relational database management system - RDBMS).

An example of the user interface for data entry and database creation „Microsoft Visual Studio - Basic 2008 Express Edition“, is shown in Figure 3

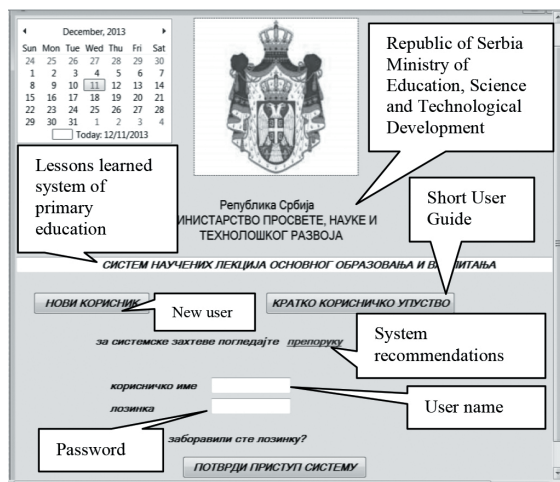


Figure 3 interface for data entry and database creation under attack „Microsoft Visual Studio - Basic 2008 Express Edition“

7. CONCLUSION

Finding patterns, trends and anomalies in data groups, and their summarizing the proposed model is one of the biggest challenges in the information age - turning data into information and information into knowledge.

The aim of this paper is focused on the development of the original, universal methods of forecasting and decision-making methods of data mining, which is confirmed in the case of building a system of lessons learned.

Increasing the level of achievement is not an end in itself, it should provide a better understanding of the real needs of both students and teachers and society in general, it is a tool for achieving better results in education.

The main problem in the existing educational policy is the lack of connection between institutions within the same system and the lack of understanding that we should learn from mistakes. Open analysis of the problems identified in the educational process will ensure that the same mistakes are not repeated in future work.

REFERENCES

- [1] Jiawei, Han, and Micheline Kamber, "Data mining: concepts and techniques," San Francisco, CA, itd: Morgan Kaufmann, 2001.
- [2] Kamber, Micheline, Jiawei Han, and Jian Pei, "Data mining: Concepts and techniques," Elsevier, 2012.
- [3] *Understanding student college choice*. Higher Education Handbook of Theory and Research, Hossler, D., Braxton, J.,& Coopersmith, G Agathon Press, New York, ., No. 5,231-288,1989.
- [4] *Merenje spremnosti zemalja za razvoj ekonomije bazirane na znanju* Ilić, V., Nedić, V., Ekonomika, Vo. 60, No 3, str. 247-257, 2014.
- [5] *Reforma produžnih škola u Minhenu i utemeljenje dualnog sistema stručnog obrazovanja*, Ilić-Rajković, A., Nastava i vaspitanje, Vol. 64, No 1, p 117-130, 2015.
- [6] *Youth unemployment and entrepreneurship*, Južnik L. Rotar, Ekonomika, Vol.60, No 4, P. 43-56, 2014.
- [7] M. Knežević, P. Veselinović, "New education policy's priorities as a function of economic development of Serbia," Ekonomika, Vol. 61, No1, p. 151-159, 2015.
- [8] B. Mikanović, "Ishodi učenja i standardi znanja u osnovnom obrazovanju," Inovacije u nastavi – časopis za savremenu nastavu, Vol. 27, No 1, p. 84-93, 2014.
- [9] B. Milošević, Z. Stojković, "Analiza strategije upisa u srednje škole i tendencije zapošljavanja u gradu Nišu," Ekonomika, Vol. 2, No 2, str. 184-190., 2013.



- [10] E. Ngai, Y. Hu, Y. Wong, Y. Chen, & X. Sun, "The application of data mining techniques in financial fraud detection," A classification framework and an academic review of literature, Decision Support Systems, 2011.
- [11] P. Nijkamp, I. Siedschlag, "Growth and Competitiveness, Dynamic Regions in the Knowledge-Based World Economy," Springer Heidelberg Dordrecht, London New York, 2011.
- [12] Z. Paszek, "Metodologija ocene kvaliteta edukacije," Škola biznisa, broj 3. str. 18-28., 2012.
- [13] V. Sučević, & M Kerić, "Element i indikatori za identifikaciju kvaliteta osnovnog obrazovanja," Sin-teze, Vol. 4, No 7, p. 19-32., 2015.
- [14] M. Tomić, V. Spasenović, & E. .Hebib, "Linerarni model srednjeg stručnog obrazovanja:primer Austrije," Nastava i vaspitanje, Vol. 64, No1, p 131-144., 2015.