Approach to the implementation of business intelligence systems to analyze the performance results of students - ICAIIT 2014

Nataša Aleksić*, Milan Erić **
* Higher Technical School of Professional Studies, Serbia
** Faculty of Engineering Sciences, Kragujevac, Serbia
e – mail :ingnaca78@yahoo.com
e-mail: ericm@kg.ac.rs

Abstract - The aim of this paper is to present the way to implement business systems intiligencije in education, which can improve the educational process, raise the quality of business processes in education and justify the high costs of implementation.

I. INTRODUCTION

With the adoption of the Bologna Declaration and the amendment of the law on higher education, higher education institutions have found themselves in a position to change their current way of doing business and it is increasingly turning to market business. That opened the door for the application of the concept of business intiligencije as an important support for the process of decision-making right and timely decisions necessary for the successful operation and management of any organization, including institutions of higher education.

When it comes to higher education institutions, the use of business intelligence, in addition to the aforementioned benefits, would bring less pressure on the students' service, a complete picture of students, improving the quality of teaching, increasing market share (more students), higher income.

Despite all the advantages that brings the concept of business intiligencije are underutilized in the field of higher education. For this reason, this work was written and, as an approach to the implementation of business intelligence systems for the analysis the performance results of students.

The introductory discussion defined the basic settings of business intelligence, where he defined the concept of business intiligencije:

- Data Warehouse
- multidimensional data analysis OLAP (*On – line Analytical Processing)* and
- detection of "knowledge" in the data (*Data Mining)*, as an integral part of modern information systems used to support decision making.

During the study, described the current state of the higher education institution. Most of the adopted criteria contain elements that require continuous collection and analysis of data on different elements of the teaching process. The necessary data include indicators of the quality of student enrollment, transition and success of the students in their studies, efficacy studies and the percentage of graduation, the structure and quality of the teaching staff, the results of the scientific work of the teaching staff, the availability and quality of natural resources and other information relaventne that enhance business activity and higher institutions, both educational and scientific - research, and business - administration.

These data represent data that change frequently, and whose analysis to obtain information that serves as a support in making business decisions. Efficient analysis of large amounts of data involves the use of adequate software support to the abundance of data has been a timely and accurate information needed to make informed business decisions.

The example is implemented using MS SQL Server 2005 and its tools for business intiligenciju, which are stored in the SQL Server Business Intelligence Development Studio.

II. BUSINESS INTELLIGENCE

Business intelligence is a term used for a set of methods and tools intended for decision support. In the field of higher education, as in many other areas of business, there is a need for implementing business systems intiligencije such systems for data analysis and support business decision making.

Business intelligence systems are applied to solving problems in the field of education, such as:

- Adapting educational materials to student needs;
- The realization of a quality system of reporting based on data collected in Learning Management Systems.

A. The basic components of business intelligence

The concept of a data warehouse

Central place of a whole strategy of business intelligence data warehouse belongs. Data warehouse (Data Warehouse) is a specially designed, analytical database that stores information collected from internal or external sources, and enabling complex, advance contingency (ad-hoc) approaches a large variety of data [1].

Data warehouse belongs to a multi - dimensional databases that are made on the basis of dimensional models. According to the creator of Inmon (WHInmon), a data warehouse is the one form of databases, which are characteristic of the four elements: [2]

- The orientation of objects (functional areas);
- Content constancy;
- The integration;
- Correspondence with time (time-defined).

Subjective oriented means that focused on specific topics, ie. Organizational activities and decision-making process.

Integration also means the use of convention names, domain restrictions, and more.

Time dependence is reflected in the data depending on the weather. Immutability means that the data warehouse is not changed permanently, but only periodically updated from operational data.

For data warehousing Most manufactures dimensional model that provides better opportunities for visualizing data, which is due to the purpose warehouses, a great advantage. It is the possibility of easy visualization of data that are inherently abstract, is the key to understanding and acceptance of the dimensional data model. Two-dimensional tabular models of transactional systems are more suitable for monitoring and management of business processes, a dimensional model of data warehouse for reporting on business processes.

Both models, and dimensional object, they are able to accept and describe the same set of data to and from them create the same set of statements or conduct the same analysis. The primary difference is in the presentation of data. Dimensional model of top-down approach, because it focuses on the specific question that the manager wanted a response, and the object model of bottom-up, showing the basic facilities required in the management or conduct of the business process.

Multidimensional data analysis – OLAP

Key to the success of the concept of business intelligence lies in the possibility that business customers for the purpose of making business decisions, provide quick and easy access to information stored in data warehouses, which are necessary for the generation of multidimensional queries, such as are usually placed in the decision-making process. One of the ways to access this information, the use of interactive technology analytical processing - OLAP (online analytical processing).

Basic characteristics possessed by any OLAP tools are: multidimensionality, drill - down, rotation, as well as several ways to view information.

The multidimensionality of the main characteristics of OLAP tools, which allows for more dimensional analysis, ie. allows users to see the determinants of business and analyzing the intersection of dimensions that describe the determinants. People's natural to observe the appearance of the business dimensions. If the phenomenon monitored in three dimensions, it is a cube, and multiple dimensions of the hypercube. Each dimension of the cube belongs to one parameter phenomenon. Each point in the cube has clearly defined the value of each of the observed dimensions.

Figure 1 shows an example of a two-dimensional OLAP cubes.

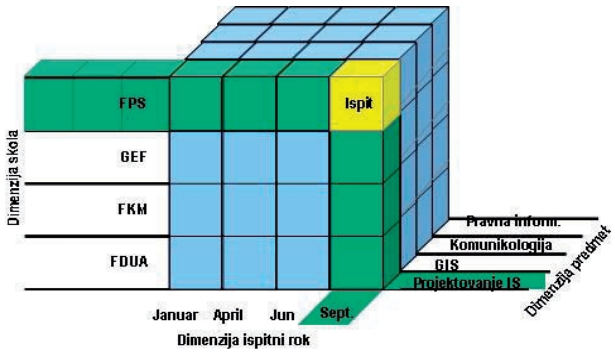

Figure 1. OLAP cube

Drill - down / up - is an analytical technique that allows the user to move around the level of data aggregation, ie. from the highest level to the corresponding detail data at the lowest level. This functionality is possible in the dimensions that have a hierarchy. In the data warehouse, must also be appropriate aggregates are defined for this hierarchy, in order for this operation was quick and efficient.

Rotation is characteristic of OLAP tools, which allows a view of the data from the reverse perspective. OLAP systems usually appear in the output section size represented by X and Y axis. Substituting values on the X and Y receives the rotated display output information.

Many different ways of displaying the output data is a very important characteristic of OLAP tools. Display output information on a graph allows better and easier comparative analysis and identification.

Knowledge discovery in data - Data mining

One of the definitions of knowledge discovery in data describing him as a non-trivial process of identifying undisputed new, potential, useful and ultimately understandable form (pattern) in the data [3].

According to another definition, Data mining is the process of discovering hidden correlations, rules and trends in the testing of large amounts of historical data (housed in a data warehouse) using a statistical method, artificial intiligencije and data visualization [4].

The data by themselves can not generate knowledge. They represent the basic form of information, which need to be managed, checked, discover and interpret in order to come to the knowledge hidden in it. Discovering the form of trends and anomalies in a large amount of stored data, poses one of the greatest challenges of the information age [5].

The basic characteristic of data mining is a multidisciplinary, bearing in mind that this technology involves elements of statistics, databases, artificial intelligence, recognizing forms and trends, access to information, knowledge acquisition and data visualization.

Data mining is currently used in two different domains: the domain of the prediction and detection of domain shape or regularities in the data.

In the field of data mining prediction algiritmi are used for forecasting the behavior of the observed entity, person or object based on the given parameters and information available.

In the domain of knowledge discovery in data, data mining algorithms allow finding the form or exceptions and deviations in the data that are not obvious and immediately observable.

Three main "pillars" of data mining are techniques of data mining, and data modeling. Some of the basic techniques and algorithms of data mining are neural networks, classification trees, techniques based on the theory of fuzzy sets, genetic algorithms, nearest neighbor technique ...

III. OBJECT-ORIENTED DEVELOPMENT OF BUSINESS INTELLIGENCE SYSTEMS

The methodology of object-oriented development of business intelligence systems consist of the following steps [6]:

Defining requirements,

• Object-oriented analysis,

• Object-oriented design,

• Implementation.

Figure 2 shows the structure of object-oriented development methodologies of business intelligence systems.

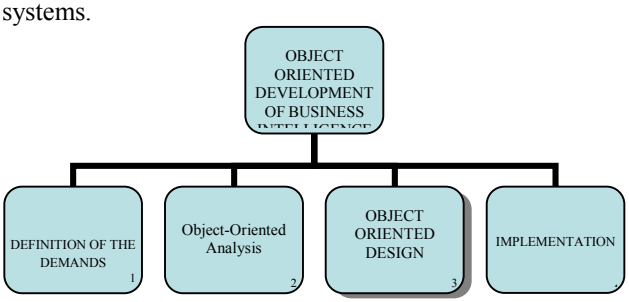

Figure 2: Schematic representation of the methodology of object-oriented development of business intelligence systems

The following points will be explained to mention the basic steps of the proposed methodology object-oriented systems development business intiligencije the example of higher education institutions. The aim of the development of the concept of business intelligence was to gain a complete picture of the student population, in order to define future strategies enrollment policy and the teaching staff as a prerequisite analysis of the performance results of students.

A. Defining requirements

The analysis requires

Higher Technical School of Professional Studies (VTŠSS) is a higher education institution, that school students of different majors. Education lasts 3 years and free choice of specialist study of 1 year. During that time shall be kept of all relaventnim data related to students. Students during the school takes an average of 20 to pass through the exam. After each examination period is evaluated examination and analysis of the effectiveness of students. Analysis of students' success during and after completion of the test period for years done the same way at a higher education institution. This paper is an attempt to get the job done better with the use of elements of the system for analytical processing (OLAP). It can be said that the decision support systems (SPO) systems which provide information to users in order to analyze situations and make decisions. In other words, it helps in making decisions that may be the strategic level, long-term, such as the analysis of the success of students, and thus increases the efficiency of the user.

Preparation of data

Data preparation is done on the basis of a predefined data sources, rules, download the data, preparation procedures and requirements. Preparation is done extracting certain - transformational tools through extraction, cleaning and transformciju data.

Figure 3 is an example of the preparation process the data.

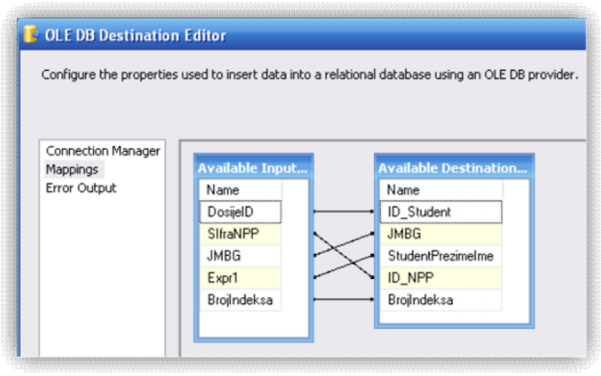

Figure 3: Preparation of data

Selection of data from existing databases

The aim of this analysis is the observation of effectiveness studies on various criteria students. To accomplish this goal, we need information about students, exams, teachers, examination terms, subjects.

The remaining part of the database is not of interest for this analysis so it can be in "full" ignore. Of the many tables selected are as follows: Student, Subject, Professor, examination and examination period.

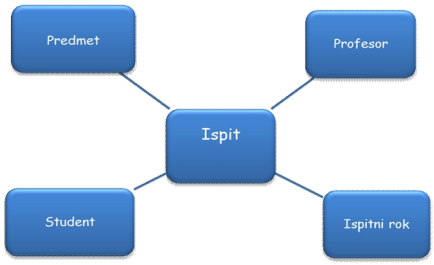

Figure 4 These star chart for each subject area and four-dimensional analysis

B. Object-oriented design

Object-oriented design is performed in three steps:

• Creating dimensional models

• Creating class diagrams

• Preparation of data for analytical processing of information

In realization of this step are used and integration services to MS SQL Server 2005th

Making dimensional model

Making dimensional model is logical design techniques designed to display data in a form that provides a high performance system for performing data analysis.

Preference dimensional model is to allow different views of the same data, as shown in Figure 5.

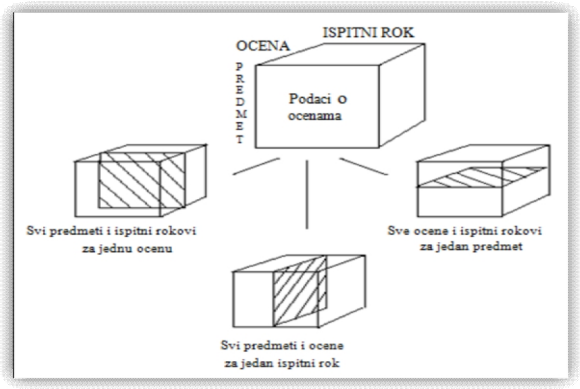

Figure 5: Different views of the same data

Dimensional model data is organized so that describe the dimensions and measures. Dimension represents an element of data that categorises each item in the data set. The primary function of dimensions is to provide filtering, grouping and labeling.

Figure 6 shows the dimensional model to analyze the success of which is based on the object-oriented analysis.

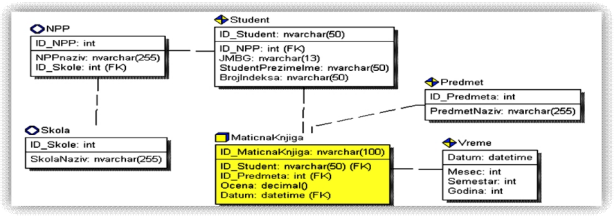

Figure 6: Dimensional model - anliza success study

IV. IMPLEMENTATION BUSINESS INTELLIGENCE SYSTEMS FOR RESULTS ANALYSIS OF STUDENT'S SUCCESS

As part of the implementation we will show results. On the model of pleural in previous phases are going to have create a data warehouse, data analysis and preparation of making the user interface.

At the beginning of the implementation phase, it is necessary to define and design the system implementation. In Figure 7 shows the structure of the system implementation..

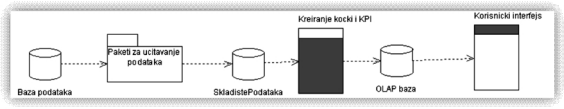

Figure 7 Design of system deployment

Creating a database is an activity that is a logical continuation of the process of modeling the data warehouse. As part of this activity should be carried out to create a physical model, generate database and perform load data into the data warehouse.

Generating a data warehouse is derived in SQL Server 2005 DBMS, using the CASE tool Rational Rose. Generating baz data is performed by using a language for defining data - Data Definition Language (DDL) for relational databases.

The physical model is translated into SQL code which is executed on the SUP and thus creates a database.

Everything that is done in the framework of the process of defining requirements, object-oriented Analysis and Object-oriented Design should allow for data analysis. The great importance of analyzing conditioned the development of various techniques for the analysis of data, some of these techniques are queries and reports generated queries, multidimensional analysis and data mining.

Figure 8 shows the physical implementation of the data warehouse

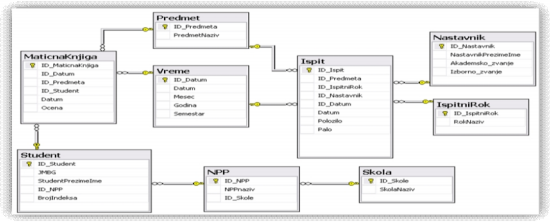

Figure 8: Physical implementation of a data warehouse in SQL Server 2005

Multidimensional analysis is a popular method of query expansion opportunities. This method of analysis replaces analysis over a large number of queries. Multidimensional analysis allows the user to perform the analysis on a large number of independent elements that make up the system to be analyzed and reviewed data that are related to each other quite complex. End users are not always interested in the same level of detail of the data so that the multidimensional analysis allows to dynamically change the level of detail.

Figure 9 shows the OLAP cube to analyze the success of students

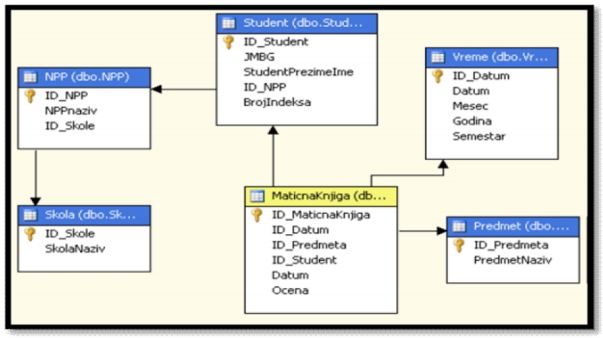

Figure 9 OLAP cube to analyze the success of students

The novelty of two-dimensional data analysis by Analysis Services SQL Server 2005 are key indicators of performances (KPI). The execution of the tasks of analysis is very useful to define key indicators of performances which are important metrics for measuring success in carrying out their business goals.

In Figure 10 is shown implemented KPI that allows no hits figures come to a conclusion by observing the status or trend.

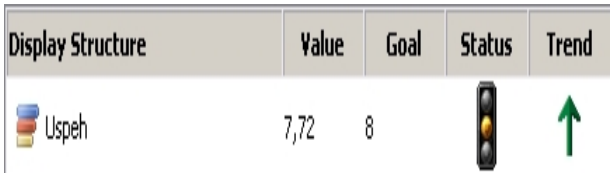

Figure 10.KPI to analyze the success of students

Analysis of data by organized in OLAP cubes can be performed using the PivotTable service that provides access to data in OLAP cubes. For the analysis of success, and with the aim of creating a system for predicting the success of students conducted a survey on a sample of 442 students from five generations, from the Higher Technical School of Vocational Studies in Kragujevac

A task that is supposed to fulfill, fulfill an artificial neural network (ANN) based on predicting the success of students at the end of undergraduate studies.

For input data, or by variables over which to neural network learns and performs prediction, we used data on the student, containing personal data and information about the success of the first study, a total of 14 different variables.

The output from the neural network for each student neural networks has created a variable representing the predicted average grade at the end of the study.

A detailed description of all of the attributes for which the ANN is presented in table 1.

Table 1. Summary of input and output attributes

variables		Description of values
The input variables	half	half of students Values: male / female
	Points on admission	The number of points earned on the entrance to enroll in university Values: 40-100 Type: real
	The study group	Study direction for which the student during enrollment in the Faculty opted Values: Information Systems and Technology and Managment Type: String
	Score in the exams of the first year	Individual marks with 11 exam the first year of basic vocational studies Value: 6-10 Type: Integer
output variables	average rating	Average rating at the end of the Faculty Value: 6-10 Type: real

In order to find the best possible quality of the model, we used six different methods of building a neural network:

- Quick: method that uses heuristic rules (rules of thumb) and the characteristics of the data in order to select the appropriate network topology.
- Dynamic: method that automatically creates an initial topology, and then adds or subtracts the hidden nodes during training.
- Multiple: a method that creates several networks with different topologies, where the exact number depends on the amount of data for training. These networks are trained parallel to, a

model with the smallest by mistake is shown as the final model.

- Prune: a method that initially creates a large network (with many hidden layers and nodes), and then removed (pruned) the weakest units in the input and hidden layers. This method is usually slow, but often gives better results than others.
- RBF (radial basis function network): uses a technique similar to k-means clustering algorithm, to particionisali data in accordance with the output variable.
- Exhaustive prune: a method that is similar to Prune method, initially created a large network (lots of hidden layers and nodes), and then removed (trimmed) the weakest unit of input and hidden layers. Parameters for Exhaustive prune method are set up to provide a very detailed search space of possible models, to find the best.

This method is the oldest, and most often gives the best results.

In the testing phase network, the by then not used the data was tested success predicting the chosen neural network model, and the results were then compared with real results. The ratio data for training and testing was 80:20.

Results comparing six different methods of neural networks are shown in Table 2, and the best values for each of the criteria of comparison are shown thickened (bold) font.

Table 2. VNM with 6 algorithms

Algorithm	*The absolute average error*	*Standardna devijacija greške*	*The linear correlation*
Quick	*0,239*	*0,239*	*0,828*
Dynamic	*0,254*	*0,29*	*0,83*
Multiple	*0,243*	*0,286*	*0,829*
Prune	*0,259*	*0,311*	*0,815*
RBFN	*0,243*	*0,307*	*0,823*
Exhaustive Prune	***0,211***	***0,262***	***0,862***

A comparative analysis of the results for all six methods of the model, the test sample (Table 2), it can be concluded that the developed ANN gives satisfactory good results. In all of the average absolute error of the prediction is in the range of 0.211 to 0.259, and linear correlation coefficient than 0, 81st .The best algorithm by all criteria can extract Exhaustive Prune, in which testing occurs when the smallest average absolute error (0.211), the lowest standard deviation (0.262), and also the largest linear correlation coefficient (0.862).

Developed multilayer ANN with Exhaustive Prune method consists of input, output and two hidden layers, which is located at 30 and 20 neurons.

View the significance of input attributes in the developed neural network model is presented in Table 3, which shows us that the seven first-ranked attributes within the developed model is considered the most significant (relative importance over 0.1).

Table 3 The importance of input variables

The input variable	*The relative importance of*
Production systems	0,159648
Mathematics 1	0,125662
The study group	0,113963
Fundamentals of Information and Communication Technology	0,106454
Fundamentals of organization	0,102788
Introduction to Information Systems	0,102465
English 2	0,100275
Mathematics 2	0,076333
English 1	0,071103
Points on admission	0,054623
Management	0,053359
Half student	0,049352
Economy	0,048522
Psychology / Sociology	0,046541

Further analysis of the significance of input variables, attributes can come to the conclusion that half of students and success in the entrance exam are from vekikog importance for predicting the success of the whole student, and to a greater extent in the impact assessment by a student on items from the first semester.

CONCLUSION

Defined are some of the possibilities of applying the techniques and tools for data analysis of student services.

In the preparation phase data were determined by the kinds of data according to the sources, performed their selection and evaluation.

The next step is to determine the data which are required for the construction of the model, or selecting data. In this step, decision which variables to keep and which to discard.

The phase transformation of data variables from the available databases are transformed into a form suitable for data mining.

Based on the available variables from the database counted attributes that are important for the solution of the problem. The development of ANN to predict the success of students have the opportunity to develop the student's pattern of behavior during the study, but the possibility of timely intervention and influence on the process of education in order to achieve success bolejeg. Also, the development of such a model provides the possibility of perceiving what aspects of the curriculum needs to be improved in order to encourage students to further work and training in specific scientific fields.

In this way, it is: better use of resources, reduce costs, make informed decisions, increase the number of enrolled students and graduates.

REFERENCES

[1] Veljović, A., relational and analytical databases ", Megatrend University, 204, p.136.

[2] Inmon, W., I., William H., „Building the data warehouse“, Wiley, 1992, str. 9-10.

[3] Oreščanin D., "Knowledge discovery and data mining," http://www.skladistenje.com/

[4] Sumathi, S., Sivananda, SN, "Introduction to Data Mining and its Applications", Springer, 2006, p. 5 .

[5] Ibidem

[6] Veljović, A., Stanojevic, Lj., "Identification of risk in the process of designing a unified banking system," Strategic Management, International Journal of strategic management and strategic management support systems, year XII, no. 3-4 / 2007, p. 17-21.