

Business Intelligence Systems for Analyzing University Students Data

Dorina Kabakchieva

University of National and World Economy, Sofia, Bulgaria

Email: dkabakchieva@unwe.bg

Abstract: *Globalization and ICT rapid development have led to strong competition between educational institutions. Advanced analytical technologies, including Business Intelligence (BI) tools, are implemented at universities worldwide for analyzing data and getting profound knowledge of the students, their individual learning characteristics and specific educational needs. This paper presents an example of BI implementation for student data analysis.*

Keywords: *Business Intelligence, university data analysis, educational data analysis.*

1. Introduction

Higher education institutions worldwide are operating today in a very dynamic and complex environment. The globalization processes and the rapid development of information and communication technologies have led to very strong competition not only between companies, but also between universities. Leading educational institutions in the USA, Australia and Western Europe, have already experienced the negative effects of those changes and have realized that they need to be proactive, introducing innovative management approaches and using progressive methods and techniques, in order to remain competitive. Universities are now focusing great efforts not only on their core competences – providing high quality education and research, but also on other activities including organizing wide marketing campaigns for attracting better students, designing and offering new attractive bachelor and master programmes in English for attracting more foreign students (valid especially for countries in which the English language is not the

native speaking language), providing opportunities to PhD and post-doctorate students for working on innovative research topics, suggesting flexible educational possibilities, etc. Modern educational institutions are entirely aware now that they are in an urgent need to analyze thoroughly the available data in order to get profound knowledge of the students, to better understand their individual learning characteristics and specific educational needs, and to adequately response to the challenging demands.

Due to the availability of powerful and affordable information technologies, the universities are collecting and keeping today large volumes of unique types of data, referring to their students, lecturers and administrative staff, to the organization and management of the educational and research processes (e.g., class and schedule information, research activities information and collection of research papers, etc.), and other managerial issues. Nowadays, the data in the educational sector is coming from two types of educational systems – traditional classroom face-to-face education and distance education. The data, generated in the two types of education, differs in many ways, including the data sources and information systems used, the data formats, and the specific goals and objectives followed.

Most Bulgarian universities are still primarily organized based on the traditional classroom educational system model. University data is collected at the admission of new students, during the organization and implementation of the educational process, and for management issues. The collected data is always stored in paper format, and some or most of it is also available in electronic format. Online information (e.g., online web pages and course content pages, multimedia databases, etc.) is also usually available. However, there is still no integrated university electronic system developed. Data is usually accumulated in different text documents and electronic tables, or it is organized in several relational databases, but there is no single data source available to provide university-wide, clean, and consistent data suitable for analysis. The educators directly interact with the students, monitor the learning process, receive a feedback, register students' attendance and could analyze the performance based on the personal observations and paper records. They rarely have more profound data to extract information that could be useful for improving the delivery of the educational material, and for taking into account the individual learning abilities and needs of the students.

In distance education, students who are separated by time and space from lecturers get an access to the educational programmes by paper-based correspondence education, videotape education, or computer-aided education. The web-based form of education, allowing students to learn via Internet, is the most popular distance education model used. The data that is generated from the different types of web-based systems (synchronous and asynchronous, collaborative and non-collaborative, etc.) is kept in various log files, containing information about the students' interactions with the system (details of timing, path, input-response, etc.). If that data is properly analyzed, very important information about the student performance could be extracted, e.g., about the preferred learning path, the most difficult problems and concepts (those which are most often visited or mistaken), etc.

However, in spite of the large volumes of data available at universities, managerial decisions are rarely taken based on it. There is a continuously growing need to transform the data into information and knowledge, considering very important organizational assets in today's knowledge-based economy and society. In response to this need, a great number of new information technologies, methods and tools have been introduced and widely used during the last 25 years, known under different terms, including Decision Support Systems, Executive Information Systems, etc., most of them being referred now as Business Intelligence (BI). Business Intelligence is a conceptual framework for decision support, combining architecture, databases (or data warehouse), analytical tools and applications, aimed at facilitating the closing of the gap between the current and desired performance of an organization as expressed in its mission, objectives and goals, and the strategy for achieving them.

However, the successful implementation of those activities requires innovative approaches. Therefore, universities now realize that to remain competitive and grab new opportunities, they need to be able to take quickly informed and competent managerial decisions. Advanced analytical technologies, including BI systems and Data Mining tools, which are already widely used by companies in all industry sectors, have also attracted the attention of educational institutions' managers. Many universities have already introduced such systems; others are currently in the process of initiating or completing such projects. The main advantages of using BI systems include integrating data from various sources and achieving "a single version of the truth" for all organization members; receiving needed reports at the right moment, at all organizational levels, and in the best possible formats; possibilities for detailed analysis of the available data through OLAP (OnLine Analytical Processing) tools; ability to find patterns and predict outcomes based on historical performance through Data Mining tools.

2. Applying Business Intelligence technologies at universities

Many universities worldwide are already using BI systems for analyzing their available data and thus supporting decision making. The cases of BI implementation in universities which are considered in this paper include higher education institutions in the USA [1-4], UK [5-8], Sweden [9] and Australia [10]. The analyzed information is taken mainly from the university websites or from case studies, describing the successful realization of the IT projects at the selected universities. The most important problem that is solved by implementing a Business Intelligence system at the analyzed universities is the integration of the data from the various sources and achieving the so called "single source of truth" which allows data consistency and thus is eliminating doubt. The importance of working with high quality data and the great efforts needed for achieving this are highlighted in most of the considered cases. Combining data from different systems in a single Business Intelligence solution facilitates cross-functional analyses and leads to higher quality decision making and efficiency improvement.

Most of the university Business Intelligence initiatives are focused in three main areas – for supporting administrative and management activities, academic issues and university relations. In some of the considered cases, the BI solutions initially deal with only some of the available data (e.g. only financial or HR data is analyzed) and later they expand to other functional areas. BI platforms support analysis, monitoring, key performance indicators and dashboards. The user interface is usually suitable for different kinds of users, simplifying the processes of requesting and analyzing data, providing facilities for personalization of reports in accordance with individual needs, online access to intuitive self-service reports, ad-hoc analysis tools, and key business performance metrics. The access to the BI platform is often provided through a web portal, to support all types of user interfaces, such as desktop computers, laptops and tablets.

Business Intelligence systems always support administrative and management activities. During university admission campaigns, student applications and offers are processed, enrollment trends are analyzed. The BI system is also used as a planning and monitoring tool, e.g. for modelling the student number plans and then monitoring during the plan period, for planning courses, for budgeting and financial planning and then monitoring the year outcomes. Monitoring is performed on Key Performance Indicators (KPIs), e.g. Progression and Achievement (the average score of student progression and achievement), Student Satisfaction, Research Assessment, Graduate Level Destination, International Population (the number of full-time international students as a proportion of the total full-time student population), to enable management to monitor their current performance against agreed strategic targets. Performance monitoring and decision making is improved also through the abilities to monitor daily operations, to apply scenario modelling, to analyze student lifecycle trends, to make forecasts, to evaluate facts and figures for an efficient ROI. This ensures that the university is kept informed of emerging trends and developments, and supports the reduction of costs related to the process of management and sharing of BI data. The BI platform is also used to transform structured data into functional information by providing capabilities for generating and distributing different kinds of reports, including universities' annual performance reports.

Academic performance is also an area that is usually widely supported by the BI solutions developed at universities. Some of the most important aspects that need accurate information for better decision making include better understanding of departmental loads, disciplines and academic outcomes; establishing areas that have performed poorly, in order to take early measures or to identify the best practice; finding the reasons for retention and monitoring progression; analyzing institutional research performance and standing; financial contributions, etc.

University relations, and particularly alumni management, are also very often supported by the university BI systems, providing accurate information and thorough analyses helping to ensure organizational sustainability and capability, university environment and engagement. Decision making is based on dashboards, scorecards and reports, allowing graphical sharing of important data with the stakeholders.

3. Needs for BI tools implementation at Bulgarian universities

Bulgarian universities have experienced many changes and are facing serious problems during the last 25 years. On one hand, the democratization processes have led to increase in the number of higher educational institutions in Bulgaria. At the moment, there are 44 universities and high schools, and 7 colleges, operating in the high education sector (Ministry of Education and Science, Registry of Higher Schools in Bulgaria, 2014) [11]. About 70% of them are publicly funded, and the other 30% are privately owned. On the other hand, the limited economic growth has led to negative demographic changes, resulting in decreased number of children born and raised in Bulgaria. Moreover, with Bulgaria becoming a member of the European Community, the opportunities for studying abroad substantially increased. These processes lead to insufficient number of potential university candidates. Many of the Bulgarian public universities have not been able to attract the usual number of students during the last years' enrollment campaigns. All these factors are seriously influencing the Bulgarian high education market, exercising high pressure especially on public institutions that have used to operate in more stable and not so severely competitive environment.

Bulgarian universities are facing new problems in their policy development today. On one hand, they build the curricula profile based on the needs of businesses and society, on the other, they should attract the right students that will be able to perform best and fulfill the university objectives. Taking into consideration the whole changing environment of educational processes within the enlarged Europe, universities are competing strongly to identify their own uniqueness and to select the most appropriate students.

University managers in Bulgaria are now confronted with the necessity of quickly taking important decisions. They could be supported in their strategy definition and fulfillment by implementing advanced analytic technologies like Business Intelligence and Data Mining, following the experiences and good practices of universities worldwide.

The main purpose of this paper is to show a simple example revealing how the implementation of BI tools for analyzing available data could contribute to taking better and informed decisions. The research is conducted with a very limited volume of data, collected by one lecturer teaching the exercise classes within a university course taught at the University of National and World Economy in Sofia, Bulgaria. The idea is to demonstrate the advantages of using Business Intelligence applications for analyzing the student performance and finding information that could be useful for improving the educational processes.

4. Research approach

The research work presented in this paper is performed in several stages – data collection and preparation, Business Intelligence software selection and application development, data analysis and findings, formulation of conclusions and recommendations.

During the first stage of the research, the data is collected and prepared for

analysis. The data comes from the personal records of a lecturer teaching exercise classes in one of the university courses taught at the University of National and World Economy. Excel software is used for the data collection. The dataset used for research purposes includes 575 records for student performance in the exercise classes of the course, in three academic years, during the period 2011-2013. Within the exercise classes, the students work with two software products of Microsoft Office package – Excel (a software for the development of electronic tables) and Access (a software for the development of relational databases), and have to pass two tests (on Excel and Access) during the semester. The student final score for the exercise classes is calculated as an average between the scores at the two tests. During the semester, the student presence in the classes is also registered. The data used for research purposes does not contain any personal information about the students. The students' names are only used for defining the student gender and then the Name field is removed from the dataset. The final set of attributes that is used for the analysis includes Gender, Year (of academic performance), Number of Absences, Excel Score, Access Score, and Final Score. The initial exploration of data reveals that the total number of 575 records refers to students that are almost evenly distributed between the two genders – 298 are females (52%) and 277 are males (48%). The records are also quite evenly distributed between the three years of performance – the dataset contains 186 records (32%) for 2011, 211 records (37%) for 2012, and 178 records (31%) for 2013, a slight prevalence of records is observed for 2012. However, these differences are not so significant and cannot be considered an obstacle for generalizing the drawn conclusions.

During the second stage of the research, software is selected for the Business Intelligence application development. Currently many Business Intelligence software tools are available at the IT market. However, most of them are commercial products which are very expensive and usually there is a limited access to them for research purposes. Fortunately, some of the leading IT vendors in the Business Intelligence sector have already started their own academic programmes and are offering academic licenses to lecturers and professors who are willing to use the software for research and educational purposes. QlikTech International [12] – a Swedish company now settled in the USA and being among the leaders in the Business Intelligence software market, is offering its Business Intelligence solution called QlikView [13] for the QlikView academic community members. QlikView is among the best Business Intelligence products available at the moment and highly appreciated by the end-users – business users working in various economic fields, as it is visible in the latest Gartner's Magic Quadrant for Business Intelligence and Analytics Platforms [14]. The software is user-friendly, easy to understand and use, providing good graphical presentation of the analyzed data and information. For all of the above mentioned reasons, QlikView is software used in the presented research for the development of the Business Intelligence application.

The developed QlikView application is organized in three dashboards on which the most important information from the data analysis is presented in the form of different graphical objects – bar and pie charts, performance indicators, tables and list boxes (used as data filters on the various attributes).

During the last stage of the research, the results of the data analysis are presented and described, conclusions are drawn and recommendations are made for future research opportunities and data collection.

5. Achieved research results

The development of the Business Intelligence application by using QlikView software is realized in two steps. At the first step, the available data in Excel format is integrated in the QlikView application. During the second step, the dashboards are designed and the graphical objects for the presentation of the data and information are developed.

The QlikView application consists of three dashboards. The first dashboard, presented in Fig. 1, provides general information about the distribution of students according to the number of their absences from the exercise classes, and according to their final performance – the achieved final score in the course exercise classes.



Fig. 1. The “General Overview” dashboard

The bar chart, presenting information about the distribution of students according to the number of absences, reveals that 85% of the students (487) have been absent in maximum three exercise classes, which is the allowed number of absences for a student per a course during a semester. Only one student has never showed up, and less than 6% have missed more than 2/3 of the exercise classes.

The data on the Final Score Distribution bar chart reveals that 34% of the students (195) have achieved excellent results (designated on the chart as 6- rounded final score value), their final score is the highest, in the interval between 5.50 and 6.00 (6.00 is considered the highest possible grade in Bulgaria, and 2.00 – the lowest one). Another third of the students (31%) have very good performance, they have achieved final scores in the interval between 4.50 and 5.49. The good performers (103) constitute 18% of all the students (575), achieving final score results in the interval between 3.50 and 4.49, and the average performers are 69 (12%). The lowest percentage fortunately is for the poor performers – they are only 29 (5%).

The distribution of the students according to the final scores achieved and based on the number of absences from the exercise classes is shown in the other two graphs – presented as a combo chart and as a line chart. As logically expected, the diagrams reveal that the students who have participated in most of the exercise classes have achieved higher results. However, the average final score for the students without any absences is 5.08 and is not excellent, which means that the participation in class is not the only requirement that will guarantee the best student performance. On the other hand, the students who have missed more than 2/3 of the exercise classes (with more than 5 absences), have achieved final scores in the interval between 3.09 and 2.25 which are actually very low results.

The second dashboard, presented in Fig. 2, is focused more on the analysis of student performance based on the student gender and the year of performance.

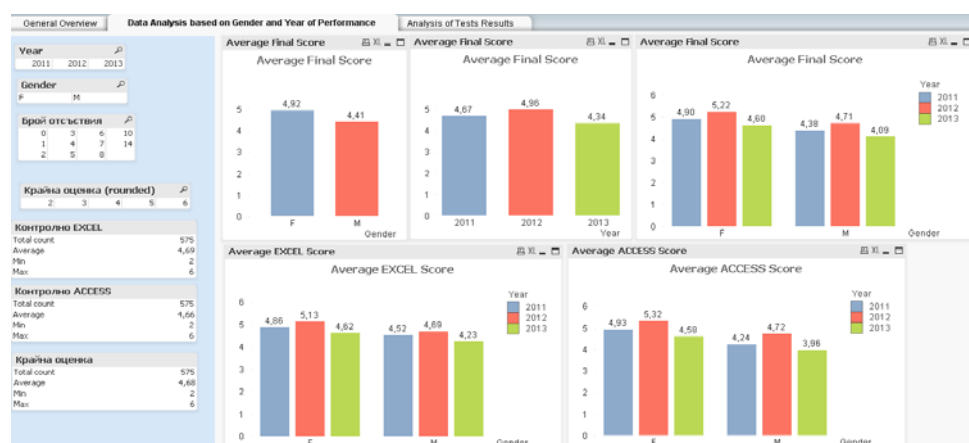


Fig. 2. The “Data analysis based on gender and year of performance” dashboard

The information presented in the bar charts reveals that the female students perform better than the male students and that is valid for the three years of performance 2011-2013. It is interesting to find out that the difference in the average final score for female and male students is always the same for the three years – 0.52 in 2011, 0.51 in 2012 and 0.51 in 2013.

The female students also perform better than the male students at both tests – Excel and Access.

The students performed best in 2012 and worst in 2013, but while the average final score achieved in 2011 and 2012 is at a very good level (i.e., above 4.50 – 4.67 and 4.96 respectively), it is at a good level in 2013 (4.34).

The third dashboard, presented in Fig. 3, provides information about the student performance in the two tests (Excel and Access), and an overview of important Key Performance Indicators (KPIs).

The information presented on the two bar charts reveals that the student performance is at similar levels for both tests (Excel and Access), since the difference for the two genders and in the three years is in the interval between 0.08 and 0.15. However, the overall analysis shows that the female students perform a little better at Access while the male students perform a little better at Excel. There is not any tendency available for the three years, but it is interesting to point out that

in 2012, when the highest average final score is achieved, the student performance is better at the Access test, while in 2011 and 2013 the students have performed better at the Excel test.



Fig. 3. The “Analysis of test results and KPIs” dashboard

The key performance indicators (KPIs) that are selected for student performance monitoring include the average values of the Final Score, Excel Score and Access Score, calculated based on the performance results of all students, and the average number of absences per student, calculated by dividing the sum of all absences to the total number of students. The KPI values for the three years (2011, 2012 and 2013) are presented in Fig. 4a-c respectively.

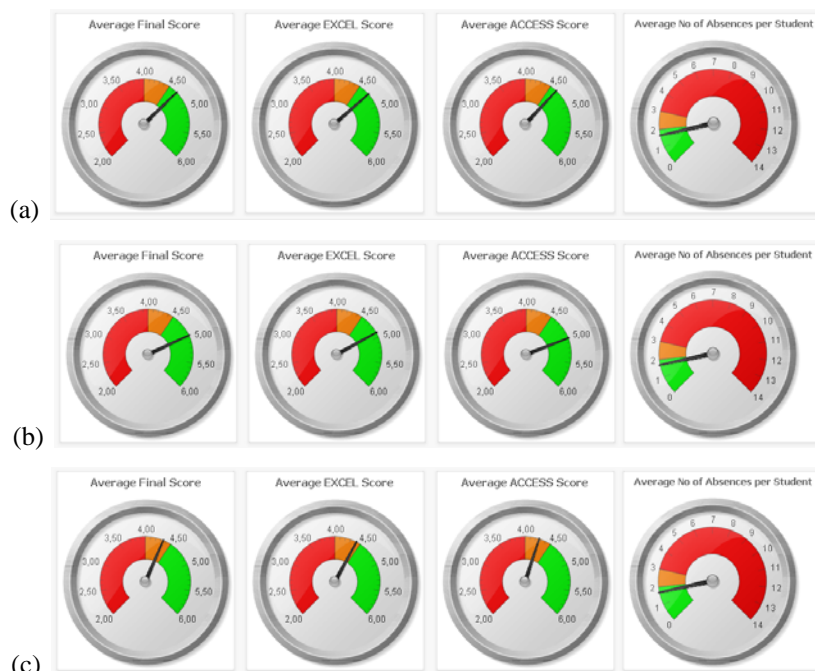


Fig. 4. The KPI values in: 2011 (a); 2012 (b); 2013 (c)

The scales for indicating the average values of the Final Score, Excel Score and Access Score, consist of three sectors. The values below 4.00 are in the red sector which means that when the indicators show such values the lecturer will be alarmed about the low performance of the students. The values between 4.00 and 4.50 are in the yellow sector and such values will attract the lecturer's attention, alarming about eventual low student performance. The values between 4.50 and 6.00 are in the green sector so that when the indicators show such values the lecturer will not be alarmed.

The scale for indicating the average number of absences per student also consists of three sectors. When the number of absences is up to 2, the indicator is in the green sector and the lecturer is not alarmed. Values between 2 and 3 are still considered acceptable and the indicator is in the yellow sector. Values above 3 are alarming and are in the red sector of the indicator.

The values of the indicators shown in Fig. 4 in the three consecutive years (2011, 2012 and 2013) reveal that the average student scores in 2011 and 2012 are in accordance with the lecturer's expectations (the indicators are in the green sectors). In 2013 the situation does not seem so good – the average values are below 4.50 (the indicators are in the yellow sectors), which is an alarm for further analysis. The average number of absences per student, for the three years, is below 2 (the indicator is in the green sector), which is acceptable because 3 is the maximum allowed number of absences per student during a semester.

The information that is made available through the Business Intelligence application development could be used for analysis of the student performance within the course exercise classes. The main findings from the conducted research reveal that about 65% (two thirds) of the students achieve very high results and only 5% fail which means that the delivered learning material is well accepted by the students and the provided support is efficient. The student participation in the exercise classes increases the chances of successfully passing the tests but this is not sufficient for achieving the highest results. The better performance of female students over the male students could probably be explained with the psychological differences between the two genders at the age of 19-20. It is a well known fact that girls are admitted at universities with higher admission scores than boys and the competition is stronger for female candidates. Girls at that age are still more ambitious and focused in their studies than boys. There is no well expressed tendency in the student performance during the three years of the considered period. However, the fact that during the last year (2013) the students performed worse is quite alarming and needs further analysis. The data analysis shows that students perform similarly at both tests (Excel and Access). This is a fact that could be used in the discussions going on among the lecturers lately, about whether to keep or exclude Access from the academic programmes of the course due to the insufficient knowledge of the first-year students about relational databases and the related difficulties in using the software tool. The usage of the key performance indicators is a good way of monitoring and controlling, providing opportunities for being proactive – establishing existing problems at an early stage, analyzing the reasons and taking adequate actions. The early awareness about the lower student performance

during the semester could be an alarm that would give an opportunity to the lecturer to provide additional support to those students that experience difficulties in understanding and working with the software tools. That could have a positive effect on the student knowledge and skills, and consequently – to the final student results.

6. Conclusions

Universities are facing serious problems in their policy development today. Taking into consideration the changing environment of the educational processes, they are competing strongly to identify their own uniqueness and to select the most appropriate students. In order to meet the challenges and the new available opportunities, the leading educational institutions worldwide have to use innovative approaches. The implementation of advanced analytical technologies, including BI systems and Data Mining tools and techniques, provides possibilities for extensive use and thorough analyses of available university data, thus ensuring more effective and efficient performance, better management and informed decision making, based on clean, accurate, secure and reliable data.

The research findings presented in the paper reveal the great potential of applying Business Intelligence tools for analyzing the available data and extracting useful information for decision making. The data used in the example is very limited and does not allow extensive analyses. Nevertheless, the achieved results are still quite interesting and provoke analytical thinking that could contribute to improving the effectiveness of the educational process. The student performance analyses could be further extended by adding new data to the initially used dataset. For example, it will be interesting to see if the best performing students are also with high performance results in the other university courses studied during the semester. Such data is available and could be extracted from the university database systems. New data could also be collected through feedback forms, distributed regularly among the students participating in the course, which will contribute to better knowing the students and their individual learning abilities.

References

1. Schaffhauser, D. Florida State U Transforms Reporting with Business Intelligence Campus Technology. 2010 (Last visited on 15 December 2014).
<http://www.tonybates.ca/2010/04/27/using-business-intelligence-tools-at-florida-state-university/>
2. Liberty University, USA. (Last visited on 16 December 2014).
<http://www.liberty.edu/informationtechnology/it-department-info/index.cfm?PID=21842>
3. Northwestern University, USA. (Last visited on 16 December 2014).
<http://www.it.northwestern.edu/business-intelligence/system.html>
4. Wesleyan University, Connecticut, USA. (Last visited on 18 December 2014).
http://www.wesleyan.edu/its/services/adminsys/business_intel.html
5. University of Central Lancashire (UCLan), UK. (Last visited on 15 December 2014).
http://www.uclan.ac.uk/about_us/bird.php

6. University of Exeter, UK. (Last visited on 16 December 2014).
<http://as.exeter.ac.uk/it/systems/businessintelligence/>
7. University of Gloucestershire – Business Intelligence Strategy. (Last visited on 18 December 2014).
<http://insight.glos.ac.uk/departments/fandp/documents/businessintelligencestrategy.pdf>
8. JISC (2011). Business Intelligence: Monitoring Performance and Planning Improvement. (Last visited on 16 December 2014).
<http://www.jisc.ac.uk/media/documents/publications/briefingpaper/2011/businessintelligence.pdf>
9. Infotool Data AB & Dimensional Insight Netherlands. 2014. A University BI-Solution: Better Insight and Control Over Operations, Easy Access for All Users, Reliable and Accessible Information. (Last visited on 18 December 2014).
http://www.infotool.se/140528_universitet_folder_lr.pdf?cms_fileid=368e75ee96405118205c58b756944989&disposition=inline
10. University of Technology, Sydney, Australia. (Last visited on 15 December 2014).
<http://www.pqu.uts.edu.au/bi-program/>
11. Ministry of Education and Science of the Republic of Bulgaria (website), (2014). Registry of Accredited Higher Schools in Bulgaria. (Last visited on 14 January 2015).
<http://www.mon.bg/?go=page&pageId=8&subpageId=167>
12. QlikTech International AB, website. 2014. Who is Qlik? (Last visited on 21 January 2015).
<http://www.qlik.com/us/company>
13. QlikTech International AB, website. 2014. QlikView Overview: Introducing the QlikView Business Discovery Platform (Last visited on 21 January 2015).
<http://www.qlik.com/us/explore/products/qlikview>
14. Gartner Inc. Magic Quadrant for Business Intelligence and Analytics Platforms. 2014 (Last visited on 14 January 2015).
<https://www.gartner.com/doc/2668318/magic-quadrant-business-intelligence-analytics>