19th International Research/Expert Conference
"Trends in the Development of Machinery and Associated Technology"
TMT 2015, Barcelona, Spain, 22-23 July 2015

DATA MINING AND PREDICTING STUDENT PERFORMANCE

Dražena Gašpar
University of Mostar,
Faculty of Economics
Matice hrvatske bb,
88 000 Mostar
Bosnia and Herzegovina

Mirela Mabić
University of Mostar,
Faculty of Economics
Matice hrvatske bb,
88 000 Mostar
Bosnia and Herzegovina

Ivica Ćorić
HERA, K.P.
Krešimira IV bb
88 000 Mostar
Bosnia and Herzegovina

ABSTRACT
This paper analyzes the possibilities of applying data mining techniques in order to improve predicting of student performance. Predicting student performance is one of the most popular applications of data mining in education. Different techniques and models are applied like neural networks, Bayesian networks, association rules, rule-based systems, regression, and correlation analysis to analyze educational data. This analysis helps in predicting student's performance i.e. to predict about his success in a course and to predict about his final grade based on features extracted from data stored in database of university.
In order to get required benefits from large data volumes stored in databases or data warehouses and to find hidden relationships between data, authors used association rules, one of the data mining techniques. Association rules are "if ... then" statements that help uncover relationships between seemingly unrelated data in a database, data warehouse or other data repository.
In this paper authors analyzed data related to students' performance on exams. By using data mining technique – association rules – authors analyzed if there is relationship between failures on exam from one subject with failure on another one.
Keywords: Data Mining, Student Performance, Association Rules

1. INTRODUCTION
Data mining methods have become very effective data analysis tools in various application domains, primarily because of their ability to deal with large volumes of structured and unstructured data and their ability to discover relevant and non-trivial information without prior knowledge [3].
Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes [5]. It was identified as one of the six emergent technologies in the NMC Horizon Report: 2013 Higher Education Edition [13]. The educational data mining was defined as "the process of converting raw data from educational systems to useful information that can be used to inform design decisions and answer research questions" [6]. It is an innovative field of research which is being implemented in education with several promising areas for data mining suggested and partially put into practice in the academic world. The main domains of data mining implementations in higher education are following [4]:
- ***Analysis and visualization of data*** is used to highlight useful information and support decision making. It can help in analyzing the students' course activities and getting a general view of a student's learning.
- ***Predicting student performance*** is the most popular application of data mining in education. Different techniques and models are applied like neural networks, Bayesian networks, rule-based systems, regression, and correlation analysis to analyze educational data. This analysis helps in predicting student's performance i.e. to predict about his success in a course.
- ***Outlier analysis*** has been used to detect and, where appropriate, remove anomalous observations from data [4].

- ***Grouping students*** means creating groups of students [15] according to their customized features, personal characteristics, etc. These clusters/groups of students can be used by the instructor/developer to build a personalized learning system which can promote effective group learning.
- - ***Enrolment management*** is frequently used in higher education to describe well-planned strategies and tactics to shape the enrolment of an institution [1] and meet established goals. Such practices often include marketing, admission policies, retention programs, and financial aid awarding.
- ***Management and generation of strategic information*** is the process of application of the IT i.e. mature strategic information system (SIS). SIS can be applied to facilitate academic and administrative activities in educational institutions. The aim is to put forward a way to understand the students' opinions, satisfactions and discontentment in the each element of the educational process [2].
- ***Target marketing*** uses data mining algorithm to generate target set which is used by marketing agent to organize promotion and marketing campaigns. The case study [7, 8] predicts the alumni pledges and helps universities to develop a cost-effective method to identify those alumni most likely to make pledges.

In this paper authors used association rules in order to discover hidden relationships between seemingly unrelated data in a database of students. Although association rules come from market basket analysis they are increasingly used in educational data mining [10, 11, 17].

Association rules capture information such as "if customers buy book X, they also buy book Y". This can be written as X→Y [9].

An association rule has two numbers that express the degree of uncertainty about the rule: support and confidence.

Let |X , Y| denotes the number of transactions that contain both X and Y. The support of that rule is the proportion of transactions that contain both X and Y: sup(X→Y) = ｜X , Y｜/n. This is also called P(X, Y), the probability that a transaction contains both X and Y. The support is symmetric: sup(X→Y) = sup(Y→X) [9].

Let |X| denotes the number of transactions that contain X. The confidence of a rule X→Y is the proportion of transactions that contain Y among the transactions that contain X: conf(X→Y) = |X , Y| /|X|. This also could be written as P(Y/X) - the probability that a transaction contains Y knowing that it contains X already. The confidence is not symmetric, usually conf(X→Y) is different from conf(Y→X), and gives its direction to an association rule [9].

Let I = {I1, , I2, ...,Ip} be a set of p items and T = {t1, , t2, ...,tn}be a set of n transactions, with each ti being a subset of I. In that case, an association rule is a rule of the form X→Y, where X and Y are disjoint subsets of I having a support and a confidence above a minimum threshold [9].

This paper presents analysis of data related to students' performance on exams. By using data mining technique – association rules – authors analyzed if there is relationship between courses which students did not pass in academic year when they first time registered for them, but they had to register for the same courses once again in the next academic year.

2. METHODOLOGY

RapidMiner Studio 6 is used for modelling association rules (Figure 1). Data is finalized in Microsoft Excel. Namely, for each course which students did not pass in academic year when they first time registered for, it was put TRUE, otherwise FALSE. Analysis was done with following measures: support>0.4 and confidence>0.6.

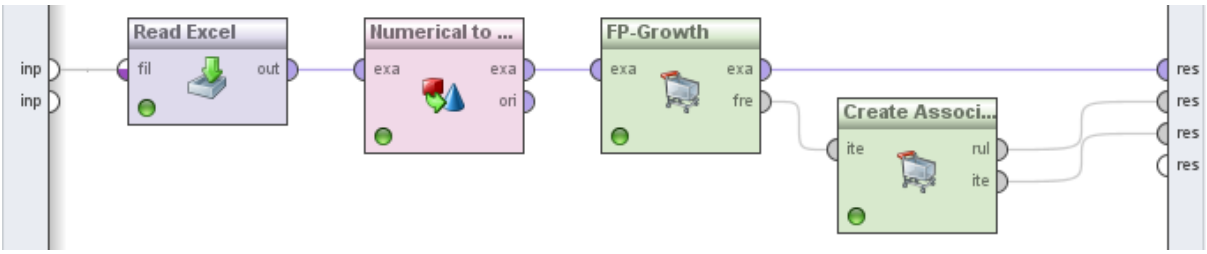

Figure 1. Modeling association rules in Rapid Miner

3. RESULTS

After data is finalized and imported in RapidMiner where modelling of association rules was done, some of obtained results were presented in Table 1 and Table 2.

Table 1. Association rules: Bachelor level – passing form the first to the second study year

Academic Year	Rule		Support	Confidence
	If	Then		
2011/12	STATISTICS	MATHEMATICS	0.415	0.800
	BUSINESS ORGANIZATIONS	MATHEMATICS	0.429	0.818
2012/13	BUSINESS ORGANIZATIONS	MATHEMATICS	0.416	0.755
	STATISTICS	MATHEMATICS	0.476	0.629
2013/14	BUSINESS ORGANIZATIONS	MATHEMATICS	0.424	0.658
	STATISTICS	MATHEMATICS	0.427	0.645
2014/15	ECONOMIC SOCIOLOGY	MATHEMATICS	0.322	0.607
	STATISTICS	MATHEMATICS	0.452	0.677
	BUSINESS ORGANIZATIONS	MATHEMATICS	0.409	0.720
All	BUSINESS ORGANIZATIONS	MATHEMATICS	0.418	0.646
	STATISTICS	MATHEMATICS	0.423	0.674

Table 2. Association rules: Bachelor level – passing form the second to the third study year

Academic Year	Rule		Support	Confidence
	If	Then		
2012/13	PUBLIC FINANCE	ACCOUNTING	0.494	0.600
2013/14	BUSINESS FINANCE	ACCOUNTING	0.244	0.600
	PUBLIC FINANCE	ACCOUNTING	0.420	0.686
	MANAGEMENT	ACCOUNTING	0.411	0.688
	ACCOUNTING	PUBLIC FINANCE	0.412	0.733
	PUBLIC FINANCE	ACCOUNTING	0.412	0.786
2014/15	ACCOUNTING	PUBLIC FINANCE	0.411	0.624
All	ACCOUNTING	PUBLIC FINANCE	0.417	0.661
	MANAGEMENT	RAČUNOVODSTVO	0.424	0.696

It is obvious from Table 1 and Table 2 that analysis was done by particular academic year and for all academic years together. Results presented in Table 1 show if students who passing from the first to the second year did not pass statistics, also did not pass mathematics. It could be said that this result was expected. As presumed, students who enrolled into first year at study of economics have different foreknowledge, especially of mathematics, because they were coming from different secondary schools. In that case, faculty could take actions to help those students in order to fulfil gaps in their knowledge. One action could be engaging better students to help their colleagues. However, it was interesting to discover that students who did not pass business organizations also did not pass mathematics. This rule is not easy to explain without additional analysis. Analyse could be done through research of students' opinion (questionnaire) about that rule and reasons behind it.

Table 2 shows rules for passing from the second to the third year of study. One association rule is that student who did not pass public finance, also did not pass accounting, and vice versa (it is symmetric rule), while other rule is that student who did not pass management, also did not pass accounting.

Since course accounting is in both rules, additional analysis of students' structure by secondary schools was made. That analysis showed that the most of students (over 60%) was coming from secondary schools that were not economic schools (gymnasiums, technical, medical schools etc.). It means that those students first time met with accounting course at the faculty. In this case, faculty could take action like initiating additional instructions to help those students to learn accounting and pass exam easier.

4. CONCLUSION

The results of research presented in this paper show that there is a significant space for use of data mining techniques in order to improve quality of the teaching process. Namely, association rules could be very useful in discovering hidden relationships among data. As it is shown in previous examples (Table 1 and Table 2), in order to understand some association rules it is necessary to do some additional analysis. But, in any case, it was shown how discovered rules could be used in taking specific actions, all with the same goal – to improve teaching and learning process at the university.
Consequently, association rules show that could be very good tool which management of university could use in order to improve overall quality, and especially quality of teaching process.

5. REFERENCES

[1] Black, J., Strategic Enrolment Management Revolution, American Association of Collegiate Registrars and Admission Officers, 2004.

[2] Bresfelean, V.P.: Data Mining Applications in Higher Education and Academic Intelligence Management. Published in: Theory and Novel Applications of Machine Learning (10. January 2009): pp. 209-228.

[3] Devine,T., Hossain,M., Harvey,E.,Baur,A.: Improving Pedagogy by Analyzing Relevance and Dependency of Course Learning Outcomes, https://pslcdatashop.web.cmu.edu/KDD2011 /papers/J-kddined2011.pdf

[4] Goyal,M., Vohra,R.: Applications of Data Mining in Higher Education, IJCSI International Journal of Computer Science Issues, Vol.9, Issue 2, No. 1, March 2012.

[5] Han, J. and Kamber, M.: Data Mining: Concepts and Techniques, 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor, 2006.

[6] Heiner, C., Baker, R., Yacef, K.: Preface. In: Workshop on Educational Data Mining at the 8 th International Conference on Intelligent Tutoring Systems (ITS 2006), Jhongli, Taiwan. 2006.

[7] Luan, J. (2002a): Data mining and knowledge management in higher education – potential applications. In Proceedings of AIR Forum, 2002, Toronto, Canada.

[8] Luan, J. (2002b): Data Mining Application in Higher Education, SPSS Executive Report, 2002.

[9] Merceron, A., Yacef,K.: Interestingness Measures for Association Rules in Educational Data, http://www.educationaldatamining.org/EDM2008/uploads/proc/6_Yacef_18.pdf

[10] Merceron, A., Yacef , K.: Educational Data Mining: a Case Study. Proceedings of Artificial Intelligence in Education (AIED2005), Amsterdam, The Netherlands, IOS Press, 2005.

[11] Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer, G., Punch, W.F.: Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA. ASEE/IEEE Frontiers in Education Conference. 2003. Boulder, CO: IEEE.

[12] Minaei-Bidgoli B., Punch W.F., Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System, GECCO 2003 Conference, Springer-Verlag, Vol 2, Chicago, USA; July 2003. pp. 2252-2263.

[13] NMC Horizon Report: 2013 Higher Education Edition, http://www.nmc.org/pdf/2013-horizon-report-HE.pdf

[14] Pimentel, E.P., Omar, N.: Towards a model for organizing and measuring knowledge upgrade in education with data mining, The 2005 IEEE International Conference on Information Reuse and Integration, Las Vegas, USA, August 15-17, 2005, p.56-60

[15] Romero, C., Ventura, S.: Educational Data Mining: A Review of the State of the Art, IEEE Transactions on Systems, Man, and Cybernetics—Part c: Applications and Reviews, vol. 40, no. 6, 2010, pp. 601-618.

[16] Vandamme, J.P., Meskens, N., Superby, J.F.: Predicting Academic Performance by Data Mining Methods, Education Economics, Volume 15, Issue 4, December 2007, p. 405-419.

[17] Wang, F.: On using Data Mining for browsing log analysis in learning environments. Data Mining in E-Learning. Series: Advances in Management Information,, Romero, C., Ventura, S., Editors, WIT press. p. 57-75, 2006.