Croatian Journal of Education
Vol.18; No.1/2016, pages: 41-70
Original research paper
Paper submitted: 8th July 2014
Paper accepted: 29th January 2015
doi: 10.15516/cje.v18i1.1471

The Use of Data Mining for Strategic Management: A Case Study on Mining Association Rules in Student Information System

Aytuğ Onan[1], Vedat Bal[2] and Burcu Yanar Bayam[3]
[1]*Department of Computer Engineering, Celal Bayar University*
[2]*Department of Management Information Systems, Celal Bayar University*
[3]*Department of Business Administration, Celal Bayar University*

Abstract

In today's competitive conditions changes in business environment and business structures make strategic management an effective form of management for business and organizations. Strategic management is a current management strategy that requires setting of the appropriate strategies, plans and applications and putting them into action in order to reach the aims and goals of organizations. The process of strategic management involves setting the company's vision, mission and objectives, determining the competitive position, and the evaluation of results obtained by strategy selection, development and application. In the application of activities related to the strategic management of business processes, the discipline of data mining, which can be defined as the process of extracting useful and meaningful patterns from large volumes of data, emerges as a viable method. In this study, strategic management and data mining disciplines and their basic concepts and applications are introduced. Apart from that, data mining methods in the context of strategic management are taken into consideration. In addition, a sample case study about the use of association rule mining algorithms in student information systems data will be presented.

Key words: *business intelligence; data mining; educational data; knowledge discovery.*

Introduction

In today's competitive environment, rapid changes in various types of business environment, such as political, economic, social or technological changes, make it crucial to consider and act strategically while managing a business or an organization. Business strategy can be defined as the management of the organization's resources and competences to reach the desired objectives by taking positive and negative external factors into account (Jeffs, 2008). Strategic management is an organizational process in which an understanding of the firm's performance is obtained by analysing external and internal environment, and the firm's conditions in terms of these environments (Nag et al., 2007). The main objectives of strategic management in the organizational context are obtaining a competitive advantage, building a sustainable competitive advantage, acting in a future-oriented way and managing the organization in an integrated manner (Barca, 2002). Strategic management is a process which requires the development of a clear vision and consequential mission statement of the firm, evaluation of strengths and weaknesses of the firm, identification of the important opportunities and threats regarding the business environment, analysis of the competition, the setting of the goals and objectives of the firm, assessment of the strategic options and selection of proper strategies, transformation of strategic plans to action plans and establishment of appropriate controls (Zimmerer et al., 2008).

Knowledge management is a vital practice for activities required in strategic management process. With the advances in the field of intelligent methods, such as expert systems, case-based reasoning, fuzzy logic, neural networks and data mining, it is possible for business organizations to extend their knowledge base by catching individual or collective knowledge (Laudon & Laudon, 2012). Expert systems, case-based reasoning and fuzzy logic methods are used for obtaining implicit knowledge, whereas neural networks and data mining are used for knowledge discovery (Laudon & Laudon, 2012).

Data mining is a non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data (Srikant & Agrawal, 1996). Data mining and knowledge discovery are important processes for obtaining competitive advantages for business organizations. Data mining methods can be used to support targeting applications, which is central to marketing management (Levin & Zahavi, 2010). With the help of data mining methods, business organizations can improve their interaction with their customers by understanding their expectations more effectively and act accordingly (Thearling, 2010). Fraud detection, financial predictive modelling, market basket analysis, customer segmentation and customer churn analysis are among other common strategic applications of data mining in business organizations (Sumathi & Sivanandam, 2006). According to Badur and Livvarçin (2006), data mining methods can be vital for strategic management by making it possible to solve strategic decision problems, providing useful patterns for competitive intelligence and by being useful for knowledge management.

Data mining in educational domain is a relatively new research direction. Educational data mining aims to explore data from educational domain with the help of designed models, tasks, methods and algorithms (Pena-Ayala, 2014). A large amount of data can be collected from educational domain for data mining applications, whereas data mining provides decision support so that the educational practice and learning material can be improved (Calders & Pechenizkiy, 2011). The application of data mining in educational setting is mainly directed towards improving learning. Student modelling, predicting the performance of students and learning outcomes, generating recommendation, analysing the behaviour of learners, communicating with stakeholders, maintaining and improving courses and comparatively analysing various forms of pedagogical support are among the main reasons why data mining should be introduced in educational practice (Bousbia & Belamri, 2014). The techniques of data mining such as information visualization, clustering, classification and association rule mining can be extremely viable tools for enhancing education. Information visualization tools provide instructors with an understanding about their students (Romero et al., 2008). Clustering can be helpful in finding students who share similar learning characteristics (Tang & McCalla, 2005). This may enable instructors to behave differently towards students, based on their personal skills and qualifications. Classification can be used to predict and model student performance (Minaei-Bidgoli & Punch, 2003). Association rule mining algorithms can be used to determine students' learning problems and offer relevant advice (Hwang et al., 2003). Strategic management in educational settings and quality assurance require activities which are knowledge-driven. In this regard, data mining plays a crucial role in improving the quality of education (Alnoukari, 2012).

Data Mining

Data mining is the process of knowledge discovery from large databases to extract useful information with the use of tools and techniques borrowed from other disciplines, such as statistics, mathematics, artificial intelligence and machine learning. Knowledge discovery process can be viewed as an iteration sequence of activities consisting of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation. Data cleaning aims to remove noisy, missing or inconsistent data from the data set; data integration aims to integrate multiple data sources; data selection aims to extract appropriate data from database; data transformation involves application of appropriate techniques so that data can be converted into a suitable form for mining task; data mining involves the application of data mining methods/algorithms to extract useful patterns; pattern evaluation aims to evaluate the interestingness of extracted patterns, while knowledge presentation involves presenting the mined knowledge to the user in a visual form (Han & Kamber, 2006).

Data mining tasks can mainly be categorized as classification, clustering, association, sequencing, regression and forecasting (Turban et al., 2005). Classification aims to predict the class of unseen data based on building a model by predefined classes, a number of attributes and a learning set (Olson & Delen, 2008). The most popular classification methods are decision tree classifiers, support vector machines (SVM), logistic regression, discriminant analysis, neural networks, Bayesian networks, K-nearest neighbour classifier, case-based reasoning, genetic algorithm and fuzzy logic based techniques (Phyu, 2009; Badur & Livvarçin, 2006). Credit scoring, bond quality rating, common stock investment category classification, common stock price and earnings performance classification, failure prediction models for non-financial firms and early warning systems for financial institutions are among the typical applications of classification methods in business and finance (Altman & Walter, 1981).

Clustering is an unsupervised learning method in which data objects are assigned to clusters such that data objects within the same cluster are as close to each other as possible, whereas data objects within assigned to different clusters are as different as possible. The assignment of data objects into clusters is handled on the basis of proximity or similarity measures between data objects (Jain & Dubes, 1988). The most important clustering methods can be classified into five categories: partitioning methods, hierarchical methods, density-based methods, grid-based methods and model-based methods. K-means, K-medoids, CLARA, CLARANS, DBSCAN and Wave Cluster are among the most well-known clustering algorithms. Market segmentation and customer segmentation are some of the representative examples of the use of clustering in business (Han & Kamber, 2006).

Association aims to identify relationships between events that occur at one time, whereas sequencing aims to identify relationships between events that occur over a time period (Turban et al., 2005). Market basket analysis, multimedia data mining, data stream mining, web mining and software bug mining are among some of the application areas of frequent pattern mining (Han et al., 2007). By applying methods for sequential analysis over a time period, sequentially purchased item sets can be obtained and this information may be useful for developing marketing plans (Badur & Livvarçin, 2006).

Regression maps the data to a prediction value. There are linear and nonlinear regression techniques. Sales predictions can be done with the aid of regression techniques. Forecasting aims to predict future trends, such as demand based on an exhibited large data set. Both regression and forecasting methods are used for estimation (Turban et al., 2005).

Methods

Association rule mining requires finding interesting associations implicit in a large amount of data. Let I be a set of attributes called items, X be an item set that is a subset of I.

Let the database D contain a set of transactions $\{T_1, T_2, \ldots, T_n\}$, where each transaction is an item set and each item set has a support or frequency value indicating statistical significance. The support s of an item set X is calculated by equation (1) given as follows (Rantzau & Schwarz, 1999):

$$s(X)=|\{T\in D|X\subseteq T\Big/\Big|D\Big|\}\Big| \tag{1}$$

An association rule is an implication X=, where X, Y and $X\cap Y$=. The confidence for rule X= is calculated by equation (2), given as follows (Rantzau & Schwarz, 1999):

$$c(X,\ Y)=s(X\cup Y)\ /\ s(X) \tag{2}$$

The confidence measure indicates the strength of a rule. A confidence threshold is used to exclude rules that are not strong enough and a support threshold is used to exclude rules whose number of transactions containing the union of antecedent and consequent part of association rule is below the specified threshold value (Rantzau & Schwarz, 1999).

Given a database D and confidence and support thresholds, association rule mining problem can be defined as the generation of all association rules X= with greater support value than the minimum support threshold and greater confidence value than the minimum confidence threshold. Association rule mining can be viewed as a two-step procedure in which finding all frequent item sets is followed by generating strong association rules from the frequent item sets (Han & Kamber, 2006). Besides support and confidence measures, some other rule evaluation measures are used in association rule mining, such as lift, rule interestingness, *J*-measure, conviction, correlation coefficients from statistics, Laplace or Gini rule induction and decision tree induction (Höppner, 2010).

Association Rule Mining Algorithms

This section presents three association rule mining algorithms utilized in this study, namely Apriori, Predictive Apriori and Tertius algorithms.

Apriori Algorithm

Apriori algorithm is one of the most popular association rule mining algorithms. It was introduced by Agrawal and Srikant (1994). Apriori algorithm finds all itemsets that have support not less than minimum support threshold based on prior knowledge. Itemsets satisfying minimum support condition are referred to as frequent itemsets. The main processing of algorithm is based on level-wise search, where k-itemsets are used to obtain $(k+1)$-itemsets (Liao, 2009). The algorithm starts with the scan of database to determine the total number of each item. The items that satisfy both minimum support and minimum confidence conditions are gathered as frequent 1-itemsets. Then, frequent 1-itemsets are used to obtain frequent 2-itemsets. In the same manner, frequent 3-itemsets are obtained from frequent 2-itemsets. The iterative process continues until no more k-itemsets can be found. An important characteristic of the algorithm is the downward closure property, which means that if an itemset is not frequent, its supersets are not frequent either (Motoda & Ohora, 2009).

Predictive Apriori Algorithm

Predictive Apriori algorithm is Apriori based algorithm for association rule mining that searches with an increasing support threshold for the best *n* rules concerning a support-based corrected confidence value. It was introduced by Scheffer (2001). Predictive Apriori algorithm aims to maximize the expected accuracy of an association rule on unobserved data. While ranking the rules, Apriori algorithm takes only confidence into account. However, predictive Apriori algorithm takes not only the confidence, but also support and predictive accuracy measures (Nahar et al., 2013).

Tertius Algorithm

Tertius algorithm is an association rule mining algorithm that searches for clauses with the highest value of confirmation evaluation. It was introduced by Flach and Lachiche (2001). Expected probability and observed probability measures are calculated in this algorithm (Nahar et al., 2013). The algorithm uses first order logic representation. The database scan depends on the number of literals in the rules. The algorithm has relatively long runtime (Arora et al., 2013).

Association Rule Mining on Educational Data

Many researchers have focused on the application of association rule mining algorithms on education data. With the use of educational association rule mining, the content which students access together, the combinations of courses students fail, students' attitudes and other useful information may be revealed. Zaiane and Luo (2001), for instance, proposed the discovery of useful patterns based on restrictions in order to aid educators in evaluating students' web-based course activities. In another study, Apriori algorithm is utilized to determine the success conditions of students based on the rules obtained on the grades received from university core courses in the first two academic years (Karabatak & İnce, 2004). Buldu and Üçgün (2010) utilized Apriori algorithm on vocational high school students' data in order to obtain the rules indicating the relation between the failed courses. Abdullah et al. (2011a) proposed a model consisting of pre-processing, mining patterns and assigning weights in order to discover highly positive association rules from students' enrolment data. Abdullah et al. (2011b) proposed a measure, called critical relative support measure, to extract efficiently least association rules to enhance current educational standards and management. In another study, association rule mining is applied to acquire useful rules from questionnaire results received from university students to observe the effects of social network sites on the students (Koç & Karabatak, 2011). Taş et al. (2013) applied Apriori algorithm to reveal the internship tendencies of students at Computer Engineering Department of Sakarya University in order to improve the efficiency of internship period and revise the internship policy of the department. In another study, students' attitudes towards selecting technical elective courses are analysed with the use of Apriori algorithm (Güngör et al., 2013).

It should be emphasized that association rule mining on education data is a promising field. Building a recommender agent for online learning activities, automatically guiding the learners' activities and intelligently generating and recommending learning materials, discovering useful relations from students' usage of information for the teacher, determining simultaneously occurring mistakes of students, obtaining a detailed feedback on e-learning process, determining navigation patterns of students and usage of these patterns in evaluation and adaptation of the course content based on students' progress are among the main viable applications of association rule mining in educational settings (Garcia et al., 2010).

Results

The data set used in this study has been obtained from student information system of Celal Bayar University. The data set contains records of 692 undergraduate students of Business Administration programme at the Faculty of Economics and Administrative Sciences. The main attributes of the data set are: class, age, gender, nationality, place of birth, family and marital status, last educational degree earned by mother, last educational degree earned by father, number of siblings, type of study (regular or evening programme), income status and success conditions of each student. Among the students involved in this study, 349 students were male and 343 students were female. The number of students enrolled in a regular study programme was 342, while the rest of the students participated in the evening programme. In this study, three well-known aforementioned association rule mining algorithms, Apriori, Predictive Apriori and Tertius algorithms, were utilized in order to extract useful association rules on the data set. In order to conduct experimental studies, WEKA (Waikato Environment for Knowledge Analysis) was used. WEKA is an open-source software written in Java, containing a collection of machine learning algorithms for data analysis and predictive modelling (Witten et al., 2011). In order to apply association rule mining algorithms, a data set is first pre-processed by unsupervised numeric to nominal conversion filter. In order to identify strong association rules regarding success conditions of the students, all students are mainly assigned into two distinct groups based on their grade point average (GPA) - as successful and unsuccessful students. Then, three association rule mining algorithms are applied to the database.

In Table 1, association rules obtained by Apriori algorithm are presented. The table illustrates the best (strong) association rules obtained by the association rule mining algorithm based on two interestingness measures: support and confidence. Here, support (which indicates the proportion of transactions in the database containing the item set) and confidence (which indicates the probability of finding the right hand side of a rule in transactions under the condition that these transactions are also in the left hand side) for the strong association rules are given.

In Table 2, association rules obtained by Predictive Apriori algorithm are presented. In Predictive Apriori algorithm, a rule is added if the expected predictive accuracy of this rule is among the *n* best rules. Hence, Table 2 includes support values and accuracy rates.

Table 1
Association rules obtained by Apriori algorithm

Rules	Support Values	Confidence
If {Age=19 AND Type_of_study=Regular_programme} ==> Class=1	73 / 73	1.0
If {Age=19 AND Success_conditions=Unsuccessful} ==> Class=1	98 / 92	0.94
If {Gender=Male AND Last_educational_degree_Father=High_school} ==> Success_conditions=Unsuccessful	90 / 84	0.93
If {Age=19} ==> Class=1	124 / 115	0.93
If {Last_educational_degree_Mother=Secondary_school_or_below AND Last_educational_degree_Father=Secondary_school_or_below AND type_of_study=Regular_programme AND Success_conditions=Unsuccessful} ==> Income_status=Low	96 / 89	0.93
If {Gender=Male AND Last_educational_degree_Father=Secondary_school_or_below AND type_of_study=Regular_programme}==>Income_status=Low	77 / 71	0.92
If {Age=19 AND Income_status=Low}==>Class=1	75 / 69	0.92
If {Class=1 AND Gender=Female AND Income_status=Low AND Last_educational_degree_Father=Secondary_school_or_below }==> Last_educational_degree_Mother=Secondary_school_or_below	88 / 80	0.91
If {Class=1 AND Last_educational_degree_Mother=Secondary_school_or_below AND Last_educational_degree_Father=Secondary_school_or_below AND type_of_study=Regular_programme}==>Income_status=Low	87 / 79	0.91
If {last_educational_degree_Mother=Secondary_school_or_below AND last_educational_degree_Father=Secondary_school_or_below AND type_of_study=Regular_programme}==>Income_status=Low	151 / 137	0.91
If {last_educational_degree_Father=Secondary_school_or_below AND type_of_study=Regular_programme AND Success_conditions=Unsuccessful}==>Income_status=Low	115 / 104	0.90
If {Age=21 AND last_educational_degree_Father=Secondary_school_or_below}==> last_educational_degree_Mother=Secondary_school_or_below	90 / 81	0.90
If {Class=1 AND Gender=Male AND last_educational_degree_Mother=Secondary_school_or_below AND last_educational_degree_Father=Secondary_school_or_below}==> Income_status=Low	90 / 81	0.90

Table 2
Association rules obtained by Predictive Apriori algorithm

Rules	Support Values	Accuracy
If {Age=19 AND Type_of_study=Regular_programme} ==> Class=1	73 / 73	0.99488
If {Class=2 AND Gender=Male AND last_educational_degree_Father=High_school} ==> Success_conditions=Unsuccessful	34 / 34	0.99411
If {Gender=Male AND last_educational_degree_MOTHER=High_school AND last_educational_degree_Father=High_school AND Type_of_study=Evening_programme} ==> Success_conditions=Unsuccessful	26 / 26	0.99338
If {Class=2 AND Age=20 AND last_educational_degree_MOTHER=Secondary_school_or_below AND last_educational_degree_Father=Secondary_school_or_below AND Success_conditions=Unsuccessful}==> Income_status=Low	25 / 25	0.99324
If {Age=19 AND Income_status=High} ==> Class=1	24 / 24	0.99309
If {Gender=Male AND place_of_birth=IZMIR AND last_educational_degree_Father=High_school AND Income_status=Low} ==> Success_conditions=Unsuccessful	23 / 23	0.99292
If {Class=2 AND Gender=Male AND last_educational_degree_Father=Secondary_school_or_below AND Type_of_study=Regular_programme} ==> Income_status=Low	21 / 21	0.99252
If {Age=21 AND Gender=Male AND last_educational_degree_MOTHER=Secondary_school_or_below AND Type_of_study=Regular_programme} ==> Income_status=Low	21 / 21	0.99252
If {Age=21 AND Gender=Female AND last_educational_degree_Father=Secondary_school_or_below AND Success_conditions=Successful} ==> last_educational_degree_MOTHER=Secondary_school_or_below	21 / 21	0.99252
If {Age=21 AND Gender=Male AND last_educational_degree_Father=High_school} ==> Success_conditions=Unsuccessful	19 / 19	0.99201
If {Age=20 AND Gender=Male AND last_educational_degree_Father=High_school AND Type_of_study=Evening_programme} ==> Success_conditions=Unsuccessful	19 / 19	0.99201
If {Class=2 AND place_of_birth=IZMIR AND last_educational_degree_Father=Secondary_school_or_below AND Success_conditions=Unsuccessful} ==> Income_status=Low	18 / 18	0.99169
If {Age=21 AND Gender=Female AND Type_of_study=Regular_programme AND Success_conditions=Successful} ==> last_educational_degree_MOTHER=Secondary_school_or_below	18 / 18	0.99169
If {Gender=Male AND last_educational_degree_Father=High_school AND Type_of_study=Evening_programme} ==> Success_conditions=Unsuccessful	52 / 51	0.99165
If {Class=2 AND Age=20 AND place_of_birth=IZMIR AND last_educational_degree_Father=Secondary_school_or_below} ==>Income_status=Low	17 / 17	0.99134
If {Age=19 AND Gender=Male AND last_educational_degree_MOTHER=Secondary_school_or_below AND last_educational_degree_Father=Secondary_school_or_below} ==> Class=1	17 / 17	0.99134

If {Age=19 AND last_educational_degree_MOTHER=Undergraduate} ==> Class=1	17 / 17	0.99092
If {Gender=Male AND place_of_birth=AYDIN} ==> Success_conditions=Unsuccessful	16 / 16	0.99092
If {Class=1 AND Age=21 AND last_educational_degree_Father=Secondary_school_or_below AND Type_of_study=Regular_programme} ==> last_educational_degree_MOTHER=Secondary_school_or_below	16 / 16	0.99092
If {Class=2 AND Age=21 AND Gender=Female Success_conditions=Successful} ==> last_educational_degree_MOTHER=Secondary_school_or_below	16 / 16	0.99092
If {Age=19 AND last_educational_degree_MOTHER=High_school AND last_educational_degree_Father=High_school AND Success_conditions=Unsuccessful} ==> Class=1	16 / 16	0.99092
If {Age=19 AND Gender=Male AND last_educational_degree_Father=High_school} ==> Class=1	15 / 15	0.99043
If {Age=23 AND last_educational_degree_Father=Secondary_school_or_below AND Income_status=Low} ==> last_educational_degree_MOTHER=Secondary_school_or_below	15 / 15	0.99043
If {Class=2 AND Age=20 AND Gender=Male AND last_educational_degree_Father=Secondary_school_or_below} ==> Income_status=Low	15 / 15	0.99043
If {Age=21 AND Gender=Male AND last_educational_degree_Father=Secondary_school_or_below AND Type_of_study=Regular_programme} ==> Income_status=Low	15 / 15	0.99043

Table 3 presents the association rules obtained by Tertius algorithm.

Table 3

Association rules obtained by Tertius algorithm

Rules
If {last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = ERZURUM OR last_educational_degree_father = Secondary_school_or_below
If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = ELAZIG OR last_educational_degree_father = Secondary_school_or_below
If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = BITLIS OR last_educational_degree_father = Secondary_school_or_below
If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = ANTALYA OR last_educational_degree_father = Secondary_school_or_below
If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = KARS OR last_educational_degree_father = Secondary_school_or_below
If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = MALATYA OR last_educational_degree_father = Secondary_school_or_below
If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = KAHRAMANMARAS OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = VAN OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = ERZINCAN OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> AGE = 28 OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = ARTVIN OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = AFYON OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = EDIRNE OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = HATAY OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = MUGLA OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> AGE = 27 OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = CANAKKALE OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> last_educational_degree_father = Secondary_school_or_below OR CLASS = 3

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = KUTAHYA OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = ANKARA OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = DIYARBAKIR OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = SIVAS OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = KASTAMONU OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = SAKARYA OR last_educational_degree_father = Secondary_school_or_below

If { last_educational_degree_mother = Secondary_school_or_below AND Income_status = Low } ==> place_of_birth = YOZGAT OR last_educational_degree_father = Secondary_school_or_below

Discussion

For Apriori algorithm, rules with confidence levels above 90% are selected. For Predictive Apriori algorithm, rules with accuracy levels above 99% are selected and for Tertius algorithm, rules with confirmation levels above 79% are selected.

For Apriori algorithm, 13 association rules obtained are presented in Table 1. As it can be observed from the results, the best rule found with the highest confidence measure value indicates that students of age 19 enrolled in regular study programme are generally in their first year of undergraduate study programme. Similarly, the second and the fourth rules have age value 19 and class 1 in their antecedent and consequent, respectively. The third rule listed in Table 1 is an indicator of the association between the gender of students, the last educational degree earned by their parents and their success condition. According to this rule, male students with fathers who graduated from high schools are expected to be unsuccessful at their courses with a relatively high confidence value. The fifth association rule shows that unsuccessful students in regular study programme who have parents with relatively low level of education are expected to have low family incomes. Similarly, five more rules have low income in their consequent parts. The eighth rule states that first-year female students with low income and with fathers who graduated from secondary school or who have a lower level of education are expected to have mothers who graduated from secondary school or lower. Similar to this rule, the twelfth rule also emphasizes an association between the parents' educational levels.

For Predictive Apriori algorithm, 25 association rules are generated. Since Apriori and Predictive Apriori algorithms are based on different metrics, the rules generated by Predictive Apriori are not exactly the same as the ones obtained by Apriori. When the association rules are grouped according to their consequents, it can be observed that the rules generated by Predictive Apriori algorithm reveal results about four conditions: class, success condition, income status and the last educational degree earned by students' mothers. The rules indicating that students are in their first year always have age value 19 in their antecedents. As it has been obtained by Apriori algorithm, the rules about students' success conditions indicate a strong association that male students having fathers who graduated from high schools are generally unsuccessful. Moreover, the rules regarding income status in their consequents are generally characterized by lower educational degree of parents in their antecedents. The last point revealed by Predictive Apriori algorithm points out that those students whose fathers have lower educational degrees are expected to have mothers who are also not well-educated. As it can be observed from the rules listed in Table 3, the rules generated by Tertius algorithm are not as rigid as the ones that are generated by either Apriori or Predictive Apriori algorithm, since these rules generally have conjunctions in their consequents. All the rules of Tertius algorithm indicate that students with low family incomes and with mothers with lower level of education are expected to have fathers who also have a lower level of education.

The main intention of the experimental study is to extract strong association rules regarding success conditions of students. On the other hand, the study is constrained by the use of demographic attributes in model formulation. Though demographic attributes can contribute to determine success conditions of students properly, these attributes are not sufficient. In order to formulate a rigid model, attributes such as individual course grades of students and students' university entrance exam grades should also be taken into consideration. However, the student information system used to extract data set is not very flexible in providing such attributes. Many of the potentially useful attributes either do not exist or they lack certain values. Hence, some interesting association rules that may be revealed otherwise cannot be obtained in this study.

Overall, a number of different rules are generated via the aforementioned association rule mining algorithms. The above results of Apriori and Predictive Apriori algorithms indicate that the parents of unsuccessful students have low family income or/and their last educational degrees conferred are relatively low (secondary school or earlier cycles of education). Hence, generally speaking, the more educated the parents, the more successful the students are at their courses. Apart from that, the rules also exhibit patterns between the family incomes and the educational levels. Again, in families with low income parents usually have relatively low educational degrees and the students from such families tend to be unsuccessful in their courses at university. Strategic management is the discipline of setting appropriate strategies, plans and applications and actions to make it possible for organizations to reach their objectives and goals. Data mining is a viable tool for providing efficient solutions to strategic management. Since education domain is a data-rich area, data mining may provide efficient solutions in terms of strategic management in education domain.

Conclusions

Data mining in educational settings aims to enhance the quality of learning process. Among the many possible application areas of educational data mining, student modelling plays a central role in achieving the desired goals of educational quality improvement. The necessity and suitability of data mining methods for strategic management of business organizations is a well-known issue. In this paper, researchers explored the use of association rule mining algorithms on educational data gathered from student information system of Celal Bayar University. The data set is mainly characterized by demographic attributes of students. Though demographic attributes on their own may not be sufficient to fully model success conditions of students, some interesting association rules are still revealed with the help of the algorithms. The obtained association rules state that there is a strong association between parents' educational level, family income and academic success. According to the results, students with low family income or parents with low educational level are expected to be unsuccessful at their academic education with a strong confidence. Hence, the improved counselling services should be provided for the students who are expected to be unsuccessful at their courses starting from their first year in undergraduate education. As has been previously mentioned, the main constraint of the study is the limitation of data set. Hence, the study should be enriched by improving the main attributes of the data set in the future.

References

Abdullah, Z., Herawan, T., Ahmad, N., & Deris, M. M. (2011a). Extracting highly positive association rules from students' enrolment data. *Procedia- Social and Behavioral Sciences,* 28, 97-101. http://dx.doi.org/10.1016/j.sbspro.2011.11.022

Abdullah, Z., Herawan, T., Ahmad, N., & Deris, M. M. (2011b). Mining significant association rules from educational data using critical relative support approach. *Procedia- Social and Behavioral Sciences,* 28, 107-111. http://dx.doi.org/10.1016/j.sbspro.2011.11.020

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. Paper presented at the 20th International Conference on Very Large Data Bases, Santiago.

Alnoukari, M. (2014). ASD-BI: A Knowledge discovery process modelling based on adaptive software development agile methodology. In R. El Sheikh, & A. Abdel (Eds.), *Business Intelligence and Agile Methodologies for Knowledge Based Organizations: Cross-Disciplinary Applications* (pp. 183-205). New York: IGI Global.

Altman, E. I., & Walter, I. (1981). *Applications of classification techniques in business, banking and finance.* Connecticut: Jai Press.

Arora, J., Bhalla, N., & Rao, S. (2013). A review on association rule mining algorithms. *International Journal of Innovative Research in Computer and Communication Engineering,* 1(5), 1246-1251.

Badur, B., & Livvarçin, Ö. (2006). *Role of data mining in strategic management*. Paper presented at the International Strategic Management Conference, Istanbul.

Barca, M. (2002). Stratejik açı: stratejik düşünme düzeyi, tarzı ve gerekliliği. In İ. Dalay, R. Coşkun, & R. Altunışık (Eds.), *Stratejik Boyutuyla Modern Yönetim Yaklaşımları* (pp. 9-26). İstanbul: Beta Publishing.

Bousbia, N., & Belamri, I. (2014). Which contribution does EDM provide to computer-based learning environments?. In A. Pena-Ayala (Ed.), *Educational Data Mining* (pp. 3-28). Berlin: Springer-Verlag. http://dx.doi.org/10.1007/978-3-319-02738-8_1

Buldu, A., & Üçgün, K. (2010). Data mining application on students' data. *Procedia- Social and Behavioral Sciences,* 2, 5251-5259. http://dx.doi.org/10.1016/j.sbspro.2010.03.855

Cai, Z., Li, Z., Kang, Z., & Liu, Y. (2009). *Computational Intelligence and Intelligent Systems*. Springer: Berlin. http://dx.doi.org/10.1007/978-3-642-04962-0_55

Calders, T., & Pechenizkiy, M. (2011). Introduction to special section on educational data mining. *ACM SIGKDD Explor.,* 13(2), 3-6. http://dx.doi.org/10.1145/2207243.2207245

Flach, P. A., & Lachiche, N. (2001). Confirmation-guided discovery of first-order rules with Tertius. *Machine Learning, 42(1-2),* 61-95. http://dx.doi.org/10.1023/A:1007656703224

Garcia, E., Romero, C., Ventura, S., de Castro, C., & Calders, T. (2010). Association Rule Mining in Learning Management Systems. In C. Romero, S. Ventura, M. Pechenizkiy, & R. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 93-107). New York: Taylor & Francis Group. http://dx.doi.org/10.1201/b10274-9

Gungor, E., Yalçın, N., & Yurtay, N. (2013). *Apriori algoritması ile teknik seçmeli ders seçim analizi*. Paper presented at Akademik Bilişim Conference, Mersin.

Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques.* (2nd ed.). Massachusetts: Morgan Kaufmann Publishers.

Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery,* 15(1), 55-86. http://dx.doi.org/10.1007/s10618-006-0059-1

Höppner, F. (2010). Association rules. In O. Maimon, & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 353-376). Berlin: Springer Verlag. http://dx.doi.org/10.1007/978-0-387-09823-4

Hwang, G. J., Hsiao, C. L., & Tseng, C. R. (2003). A computer-assisted approach to diagnosing student learning problems in science courses. *Journal of Information Science and Engineering*, 19, 229-248.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data.* Upper Saddle River: Prentice Hall.

Jeffs, C. (2008). *Strategic management.* (1st ed.). Sage: London.

Karabatak, M., & İnce, M. C. (2004). *Apriori algoritması ile öğrenci başarısı analizi*. Paper presented at Eleco'2004-Elektrik-Elektronik ve Bilgisayar Mühendisleri Sempozyumu, Bursa.

Laudon, K. C., & Laudon, J. P. (2011). *Management information systems.* (12th ed.). London: Prentice Hall.

Levin, N., & Zahavi, J. (2010). Data mining for target marketing. In O. Maimon, & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 1261-1299). Berlin: Springer Verlag. http://dx.doi.org/10.1007/978-0-387-09823-4

Liao, B. (2009). An improved algorithm of Apriori. *Communications in Computer and Information Sciences,* 51, 427-432. http://dx.doi.org/10.1007/978-3-642-04962-049

Minaei-Bidgoli, B., & Punch, W. (2003). *Using genetic algorithms for data mining optimization in an educational web-based system* (pp. 2252–2263). Paper presented at the Genetic and evolutionary computation conference, Chicago, USA. http://dx.doi.org/10.1007/3-540-45110-2_119

Motoda, H., & Ohara, K. (2009). Apriori. In X. Wu, & V. Kumar (Eds.), *The Top Ten Algorithms in Data Mining* (pp. 61-93). NY: Taylor and Francis Publishing, (Chapter 4) http://dx.doi.org/10.1007/s10115-007-0114-2

Nag, R., Hambrick, D. C., & Chen, M. J. (2007). What is strategic management, really? Inductive derivation of a consensus definition of the field. *Strategic Management Journal,* 28, 935-955. http://dx.doi.org/10.1002/smj.615

Nahar, J., Imam, T., Tickle, K. S., & Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications,* 40, 1086-1093. http://dx.doi.org/10.1016/j.eswa.2012.08.028

Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques.* Berlin: Springer-Verlag. http://dx.doi.org/10.1007/978-3-540-76917-0

Pena-Ayala, A. (2014). Educational data mining: a survey and a data mining-based analysis of recent works. *Expert Systems with Applications,* 41(4), 1432-1462. http://dx.doi.org/10.1016/j.eswa.2013.08.042

Phyu, T. N. (2009). Survey of classification techniques in data mining. *Proceedings of the International MultiConference of Engineers and Computer Scientistis,* 1, 727-731.

Rantzau, R., & Schwarz, H. (1999). *A* multi-tier architecture for high-performance data mining. In A.P. Buchman (Ed.), *Datenbanksysteme in Büro, Technik und Wissenschaft Informatik aktuell* (pp. 151-163). Berlin: Springer-Verlag. http://dx.doi.org/10.1007/978-3-642-60119-4_10

Romero, C., Ventura, S., & Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education,* 51(1), 368-384. http://dx.doi.org/10.1016/j.compedu.2007.05.016

Scheffer, T. (2001). *Finding association rules that trade support optimally against confidence.* Paper presented at the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases, Freiburg. http://dx.doi.org/10.1007/3-540-44794-6_35

Srikant, R., & Agrawal, R. (1996). *Mining sequential patterns: generalizations and performance improvements.* Paper presented at the 5th International Conference on Extending Database Technology, France. http://dx.doi.org/10.1007/bfb0014140

Sumathi, S., & Sivanandam, S. N. (2006). *Introduction to data mining and its applications.* Berlin: Springer-Verlag. http://dx.doi.org/10.1007/978-3-540-34351-6_2

Tang, T., & McCalla, G. (2005). Smart recommendation for an evolving e-learning system. *International Journal on E-Learning*, 4(1), 105-129.

Taş, M., Adak, M. F., & Yurtay, N. (2013). *Birliktelik kurallarının çıkarılması ve staj eğiliminin belirlenmesi.* Paper presented at ISITES 2013-The First International Symposium on Innovative Technologies in Engineering and Science, Sakarya.

Thearling, K. (2010). Data mining for CRM. In O. Maimon, L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 1249-1252). Berlin: Springer Verlag. http://dx.doi.org/10.1007/978-0-387-09823-4

Turban, E., Aronson, J. E., & Liang, T. P. (2004). *Decision support systems and intelligent systems.* (7th ed.). Upper Saddle River: Prentice-Hall.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques.* (3rd ed.). San Francisco: Morgan Kaufmann.

Zaiane, O., & Luo, J. (2001). *Web usage mining for a better web-based learning environment*. Paper presented at Conference on Advanced Technology for education, Alberta.

Zimmerer, T. W., & Scarborough, N. M. (2008). *Essentials of entrepreneurship and small business management.* (5th ed.). London: Prentice Hall.

Aytuğ Onan
Department of Computer Engineering,
Faculty of Engineering,
Celal Bayar University, 45140, Manisa, Turkey
aytug.onan@cbu.edu.tr

Vedat Bal
Department of Management Information Systems,
Faculty of Business Administration,
Celal Bayar University, 45140, Manisa, Turkey
vedat.bal@cbu.edu.tr

Burcu Yanar Bayam
Division of Business Administration,
Graduate School of Social Sciences,
Celal Bayar University, 45140, Manisa,Turkey
burcu.yanar@cbu.edu.tr

Upotreba rudarenja podataka u strateškom menadžmentu: analiza slučaja upotrebe pravila pridruživanja rudarenja podataka u informacijskom sustavu podataka o studentima

Sažetak

U današnje vrijeme kada je konkurencija svugdje jaka, promjene u poslovnom okruženju i poslovnim strukturama čine strateški menadžment učinkovitim oblikom menadžmenta u kompanijama i organizacijama. Strateški menadžment je suvremena strategija koja zahtijeva postavljanje odgovarajućih strategija, planova i programa, kao i njihovo provođenje da bi organizacije postigle svoje ciljeve. Proces strateškog menadžmenta podrazumijeva određivanje vizije, misije i ciljeva kompanije, određivanje njezine konkurentnosti i evaluaciju rezultata dobivenu odabirom, razvojem i primjenom strategije. Primjenom aktivnosti povezanih sa strateškim menadžmentom u poslovnim procesima disciplina rudarenja podataka pokazala se jako uspješnom metodom. Ona se može definirati kao proces izdvajanja korisnih i smislenih uzoraka iz veličinom golemih podataka. U ovom istraživanju prikazuju se discipline strateškog menadžmenta i rudarenja podataka, kao i njihovi osnovni pojmovi i primjena. Osim toga, u obzir su uzete metode rudarenja podataka u kontekstu strateškog menadžmenta. Uz to će se još prezentirati i primjer analize slučaja o upotrebi algoritama pravila pridruživanja u rudarenju podataka u sustavu podataka o studentima.

Ključne riječi: *otkrivanje znanja; poslovna inteligencija; podaci o obrazovanju; rudarenje podataka.*

Uvod

U današnjem konkurentnom okruženju brze promjene u raznim vrstama poslovnog okruženja, poput političkih, ekonomskih, društvenih ili tehnoloških promjena, zahtijevaju strateško razmišljanje i djelovanje pri upravljanju kompanijom ili organizacijom. Poslovna strategija može se definirati kao upravljanje resursima i kompetencijama organizacije da bi se postigli željeni ciljevi, tako što se u obzir uzimaju i pozitivni i negativni vanjski čimbenici (Jeffs, 2008). Strateški menadžment je organizacijski proces u kojem se razumijevanje načina na koji tvrtka radi temelji na analizi vanjskog i unutarnjeg okruženja te položaja tvrtke s obzirom na te uvjete (Nag i sur., 2007). Glavni ciljevi strateškog menadžmenta u organizacijskom kontekstu su stjecanje konkurentske prednosti, stvaranje održive konkurentske prednosti, djelovanje okrenuto budućnosti i upravljanje organizacijom na integrirani način (Barca, 2002). Strateški menadžment je proces koji od tvrtke zahtijeva razvijanje jasne vizije i misije koja proizlazi iz nje, evaluaciju snaga i slabosti tvrtke, prepoznavanje važnih prilika i prijetnji povezanih s poslovnim okruženjem, analizu konkurencije, postavljanje zadataka i ciljeva tvrtke, procjenu strateških opcija i odabir odgovarajućih strategija, transformaciju strateških planova u akcijske planove i uspostavljanje odgovarajućih načina kontrole (Zimmerer i sur., 2008).

Upravljanje znanjem neizbježno je kod svih aktivnosti koje zahtijeva proces strateškog menadžmenta. Zahvaljujući napretku ostvarenom u polju inteligentnih metoda poput ekspertnih sustava, zaključivanja na temelju slučaja, neizrazite logike, neuronskih mreža i rudarenja podataka, poslovne organizacije mogu proširivati svoju bazu znanja tako što imaju mogućnost evidentirati znanje pojedinca ili kolektivno znanje (Laudon i Laudon, 2012). Metode ekspertnih sustava, zaključivanja na temelju slučaja i neizrazite logike koriste se za dobivanje implicitnog znanja, a neuronske se mreže i rudarenje podataka koriste za otkrivanje znanja (Laudon i Laudon, 2012).

Rudarenje podataka je netrivijalan proces prepoznavanja valjanih, novih, potencijalno korisnih i potpuno razumljivih uzoraka u podacima (Srikant i Agrawal, 1996). Rudarenje podataka i otkrivanje znanja poslovnim su organizacijama važni procesi za dobivanje konkurentske prednosti. Metode rudarenja podataka mogu se koristiti kao podrška ciljanim programima, što je ključno u marketinškom menadžmentu (Levin i Zahavi, 2010). Uz pomoć metoda rudarenja podataka poslovne organizacije mogu poboljšati svoju interakciju s klijentima tako što mogu bolje razumjeti njihova očekivanja i djelovati na odgovarajući način (Thearling, 2010). Otkrivanje prijevare, prediktivni financijski model, analiza potrošačke košarice, segmentacija kupaca i *churn* analiza kupaca još su neke od čestih strateških primjena rudarenja podataka u poslovnim organizacijama (Sumathi i Sivanandam, 2006). Prema Baduru i Livvarçinu (2006), metode rudarenja podataka mogu biti od presudne važnosti za strateški menadžment jer omogućavaju rješavanje problema strateškog odlučivanja, pružaju korisne uzorke konkurentske inteligencije i korisne su u upravljanju znanjem.

Rudarenje podataka u obrazovnoj domeni relativno je nov smjer u istraživanju. Rudarenje podataka u području obrazovanja ima za cilj istražiti podatke u obrazovnoj domeni uz pomoć postojećih modela, zadataka, metoda i algoritama (Pena-Ayala, 2014). Iz obrazovne domene može se prikupiti velika količina podataka za programe rudarenja podataka, a samo rudarenje podataka pruža podršku odlučivanju i omogućava poboljšanje obrazovne prakse i nastavnih materijala (Calders i Pechenizkiy, 2011). Primjena rudarenja podataka u obrazovnom okruženju ponajprije ima za cilj poboljšati proces učenja. Modeliranje studenata, predviđanje njihova uspjeha i obrazovnih ishoda, izrada preporuka, analiza ponašanja studenata, komunikacija sa sudionicima, održavanje i razvijanje kolegija, kao i komparativna analiza različitih oblika pedagoške podrške među glavnim su razlozima zašto bi rudarenje podataka trebalo uvesti u obrazovnu praksu (Bousbia i Belamri, 2014). Tehnike rudarenja podataka kao što su vizualizacija informacija, grupiranje podataka, njihova klasifikacija i pravila pridruživanja mogu biti vrlo uspješni alati za poboljšanje standarda obrazovanja. Alati za vizualizaciju informacija pomažu nastavnicima da bolje razumiju svoje studente (Romero i sur., 2008). Grupiranje podataka može biti od velike pomoći pri prepoznavanju studenata koji imaju slične karakteristike učenja (Tang i McCalla, 2005). To nastavnicima može dati priliku da se drugačije postave prema studentima, u skladu s njihovim osobnim vještinama i kvalifikacijama. Klasifikacija se može koristiti pri predviđanju i modeliranju uspjeha studenata (Minaei-Bidgoli i Punch, 2003). Algoritmi pravila pridruživanja mogu se koristiti za prepoznavanje problema u učenju koje studenti mogu imati te im se tako pravodobno može dati koristan savjet (Hwang i sur., 2003). Strateški menadžment u obrazovnom okruženju i osiguravanje kvalitete zahtijevaju aktivnosti koje se temelje na znanju. S obzirom na to, rudarenje podataka ima glavnu ulogu u poboljšanju kvalitete obrazovanja (Alnoukari, 2012).

Rudarenje podataka

Rudarenje podataka je proces otkrivanja znanja iz velikih baza podataka da bi se izdvojile korisne informacije. Pri tome se koriste alati i tehnike posuđene iz drugih disciplina, kao što su statistika, matematika, umjetna inteligencija i strojno učenje. Proces otkrivanja znanja može se gledati kao uzastopni niz aktivnosti koji se sastoji od čišćenja podataka, integriranja podataka, odabira podataka, transformacije podataka, rudarenja podataka, evaluacije obrazaca i prezentacije znanja. Čišćenje podataka ima za cilj ukloniti bučne, nepotpune ili nedosljedne podatke iz skupa podataka; integriranje podataka ima za cilj integrirati višestruke izvore podataka; odabir podataka ima za cilj izdvojiti odgovarajuće podatke iz baze podataka; transformacija podataka uključuje primjenu odgovarajućih tehnika da bi se podaci mogli pretvoriti u odgovarajući oblik za rudarenje; rudarenje podataka obuhvaća primjenu metoda/algoritama rudarenja podataka da bi se izdvojili korisni uzorci; evaluacijom podataka procjenjuje se zanimljivost izdvojenih uzoraka, a prezentacija znanja odnosi se na prezentaciju znanja dobivenog rudarenjem podataka korisniku u vizualnom obliku (Han i Kamber, 2006).

Zadaci rudarenja podataka mogu se uglavnom podijeliti na klasifikaciju, grupiranje, pridruživanje, stavljanje u niz, regresiju i predviđanje (Turban i sur., 2005). Klasifikacija se koristi za predviđanje skupa neviđenih podataka na osnovi izrade modela s pomoću prije definiranih skupova, brojnih atributa i skupa učenja (Olson i Delen, 2008). Najpopularnije metode klasifikacije su klasifikatori stabla odlučivanja, strojevi s potpornim vektorima (SVM), logistička regresija, diskriminantna analiza, neuronske mreže, Bayesove mreže, metoda K-najbližih susjeda, zaključivanje na temelju slučaja, genetski algoritam i tehnike utemeljene na neizrazitoj logici (Phyu, 2009; Badur i Livvarçin, 2006). Kreditni skoring, rangiranje obveznica prema kvaliteti, klasificiranje kategorije ulaganja u obične dionice, klasificiranje cijene običnih dionica i zarade, modeli predviđanja stečaja za nefinancijska poduzeća i sustavi ranog upozorenja za financijske institucije neke su od tipičnih primjena metoda klasifikacije u poslovnom i financijskom svijetu (Altman i Walter, 1981).

Klastering je metoda nenadgledanog učenja u kojoj se podaci pridružuju klasterima tako da su svi podaci unutar istog klastera jedan drugome blizu što je više moguće. Raspoređivanje podataka u klastere provodi se na temelju vrijednosti blizine ili sličnosti između podataka (Jain i Dubes, 1988). Postoji 5 kategorija najvažnijih klaster metoda: metode participiranja, metode hijerarhije, metode utemeljene na gustoći, *grid* metode i metode utemeljene na modelu. K-srednja vrijednost, K-medoidi, CLARA, CLARANS, DBSCAN i WaveCluster neki su od najpoznatijih algoritama. Segmentacija tržišta i segmentacija kupaca neki su od reprezentativnijih primjera upotrebe klasteringa u tvrtkama (Han i Kamber, 2006).

Pridruživanje ima za cilj prepoznati veze između događaja koji se događaju jednom, a sekvencioniranje ima za cilj prepoznati veze između događaja koji se događaju tijekom nekog razoblja (Turban i sur., 2005). Analiza potrošačke košarice, rudarenje multimedijskih podataka, rudarenje tokova podataka, rudarenje internetskih podataka i rudarenje po softverskim bugovima neki su primjeri područja u kojima se primjenjuje rudarenje podataka prema često korištenim uzorcima (Han i sur., 2007). Primjenom metode sekvencijalne analize tijekom određenog razdoblja mogu se dobiti podaci o uzastopno kupljenim artiklima pa se takvi podaci mogu koristiti za razvijanje marketinških planova (Badur i Livvarçin, 2006).

Regresija preslikava podatke na predviđene vrijednosti, a postoje linearne i nelinearne regresijske tehnike. Predviđanja o očekivanoj prodaju mogu se izraditi uz pomoć regresijskih tehnika. Prognoziranjem se na temelju velikog prikazanog broja podataka pokušavaju predvidjeti budući trendovi, kao što je potražnja. I regresija i metode predviđanja koriste se za procjenjivanje (Turban i sur., 2005).

Metode

Pravilo pridruživanja u rudarenju podataka zahtijeva pronalaženje zanimljivih veza u velikoj količini podataka. Neka je I skup atributa koje nazivamo elementima, a X je skup elemenata koji je podskup skupa I.

Neka baza podataka D sadrži skup transakcija $\{T_1, T_2, \ldots, T_n\}$, pri čemu je svaka transakcija skup elemenata, a svaki skup elemenata ima potporu ili vrijednost učestalosti pojavljivanja koja pokazuje statističku značajnost. Potpora s skupa elemenata X izračunava se s pomoću sljedeće jednadžbe (1) (Rantzau i Schwarz, 1999):

$$s(X) = | \{ T \in D | X \subseteq T \Big| / \Big| D \Big| \} \Big| \qquad (1)$$

Pravilo pridruživanja je implikacija *X*=, gdje je, *X, Yi X i X* $\cap$*Y*. Pouzdanost pravila *X*= izračunava se jednadžbom (2), koja slijedi (Rantzau i Schwarz, 1999):

$$c(X, Y)=s(X \cup Y) \,/\, s(X) \qquad (2)$$

Stupanj pouzdanosti pokazuje snagu pravila. Prag pouzdanosti koristi se za isključivanje pravila koja nisu dovoljno jaka, a prag potpore koristi se za isključivanje pravila čiji je broj transakcija koje sadržavaju i prethodnik i sljedbenik pravila pridruživanja ispod određene vrijednosti praga (Rantzau i Schwarz, 1999).

Uz danu bazu podataka *D,* pragove pouzdanosti i potpore, problem pravila pridruživanja u rudarenju podataka može se definirati kao stvaranje svih pravila pridruživanja $X \Rrightarrow Y$ čija je vrijednost potpore veća od minimalnog praga potpore, a vrijednost pouzdanosti im je veća od minimalnog praga pouzdanosti. Pravilo pridruživanja može se gledati kao postupak koji se sastoji od dva koraka, u kojem nakon pronalaženja svih frekventnih skupova elemenata slijedi stvaranje jakih pravila pridruživanja na temelju tog frekventnog skupa elemenata (Han i Kamber, 2006). Osim mjera potpore i pouzdanosti u pravilima pridruživanja koriste se i neke druge mjere u procjenjivanju pravila, kao što su: lift, pravilo zanimljivosti, mjera J, uvjerljivost, koeficijenti korelacije iz statistike, Laplaceovo ili Ginijevo pravilo zaključivanja, zaključivanje s pomoću stabla odlučivanja (Höppner, 2010).

Algoritmi pravila pridruživanja u rudarenju podataka

Ovaj dio rada prikazuje 3 algoritma pravila pridruživanja u rudarenju podataka, a koji su se koristili u ovom radu: Apriori, Predictive Apriori i Tertius algoritam.

Apriori algoritam

Apriori algoritam je jedan od najpopularnijih algoritama pravila pridruživanja u rudarenju podataka. Prvi su se njime koristili Agrawal i Srikant (1994). Apriori algoritam pronalazi skupove podataka čija potpora nije manja od minimalnog praga potpore određenog na temelju prethodnog znanja. Skupovi podataka koji udovoljavaju kriteriju minimalne potpore nazivaju se frekventni skupovi podataka. Glavni rad algoritma temelji se na pretraživanju po razinama, pri čemu se *k*-skupovi podataka koriste za dobivanje *(k+1)* skupova podataka (Liao, 2009). Algoritam započinje skeniranjem baze podataka da bi se odredio ukupan broj svakog podatka. Podaci koji zadovoljavaju i uvjet minimalne potpore i uvjet minimalne pouzdanosti grupiraju se u frekventne 1-skupove podataka. Nakon toga se frekventni 1-skupovi podataka koriste za dobivanje frekventnih 2-skupova podataka. Na isti se način od 2-skupova podataka dobivaju 3-skupovi podataka. Proces se ponavlja dok se ne dođe do točke kada se više ne može pronaći nijedan *k*-skup podataka. Jedna važna karakteristika algoritma je svojstvo zatvaranja prema dolje, što znači da ako neki skup podataka nije frekventan, onda nisu frekventni ni njegovi nadskupovi (Motoda i Ohora, 2009).

Prediktivni Apriori algoritam

Prediktivni Apriori algoritam temelji se na Apriori algoritmu za pravilo pridruživanja u rudarenju podataka koji s rastućim pragom potpore pretražuje najbolja *n* pravila koja imaju veze s ispravljenom vrijednošću pouzdanosti, a koja se temelji na potpori. Taj algoritam je u literaturu uveo Scheffer (2001). Prediktivni Apriori algoritam ima za cilj maksimalno povećati očekivanu točnost pravila pridruživanja na nepromatranim podacima. Pri rangiranju pravila Apriori algoritam uzima u obzir samo pouzdanost. Međutim, Prediktivni Apriori algoritam ne uzima u obzir samo pouzdanost, već i potporu i prediktivne mjere točnosti (Nahar i sur., 2013).

Tertius algoritam

Tertius algoritam je algoritam pravila pridruživanja u rudarenju podataka koji traži rečenice s najvećom vrijednošću potvrde evaluacije. Prvi su se njime koristili Flach i Lachiche (2001). Mjere očekivane vjerojatnosti i promatrane vjerojatnosti računaju se s pomoću tog algoritma (Nahar i sur., 2013). Algoritam se koristi logikom prvog reda. Skeniranje baze podataka ovisi o broju literala u pravilima. Taj algoritam ima relativno dugo vrijeme izvršavanja (Arora i sur., 2013).

Primjena pravila pridruživanja u rudarenju podataka na podatke o obrazovanju

Mnogi su se stručnjaci bavili primjenom algoritama pravila pridruživanja u rudarenju podatka na podatke o obrazovanju. Upotrebom pravila pridruživanja u rudarenju podataka u obrazovanju mogu se dobiti informacije o sadržajima kojima se koriste svi studenti, o kombinacijama kolegija na kojima su studenti neuspješni, o stavovima studenata i druge korisne informacije. Zaiane i Luo (2001) su, na primjer, pokazali način otkrivanja korisnih uzoraka na temelju ograničenja, da bi se pomoglo ocjenjivačima pri ocjenjivanju rada studenata u kolegijima u kojima rade aktivnosti putem interneta. U drugom istraživanju koristio se Apriori algoritam da bi se odredili uvjeti uspjeha studenata na temelju pravila dobivenih s pomoću ocjena iz obveznih sveučilišnih kolegija tijekom prve dvije akademske godine (Karabatak i İnce, 2004). Buldu i Üçgün (2010) su primijenili Apriori algoritam na podatke o učenicima iz srednje strukovne škole da bi dobili pravilo koje upućuje na vezu između dva kolegija u kojima su studenti neuspješni. Abdullah i sur. (2011a) su predstavili model koji se sastoji od prethodne obrade, uzoraka u rudarenju i dodjeljivanja opterećenja da bi se došlo do jako pozitivnih pravila pridruživanja na temelju podataka o upisu studenata. Abdullah i sur. (2011b) su predložili mjeru nazvanu mjera kritične relativne potpore da bi uspješno izdvojili najmanja pravila pridruživanja za poboljšanje standarda i menadžmenta u području obrazovanja. U jednom drugom istraživanju koristilo se pravilo pridruživanja da bi se uočila korisna pravila na temelju rezultata ankete koju su ispunili studenti sa sveučilišta, a s ciljem ispitivanja utjecaja društvenih mreža na studente (Koç and Karabatak, 2011). Tas i sur. (2013) su se koristili Apriori algoritmom za utvrđivanje sklonosti studenata informatike na Sveučilištu u Sakaryji prema stažiranju i praktičnom radu. Rezultati su pomogli u poboljšanju kvalitete stažiranja i doradi politike stažiranja Odsjeka za informatiku. U još jednom istraživanju su se s pomoću Apriori algoritma analizirali stavovi studenata o odabiru izbornih tehničkih kolegija (Güngör i sur., 2013).

Trebalo bi naglasiti da je primjena pravila pridruživanja na podatke o obrazovanju područje s velikim potencijalom. Izrada agenta preporuke za *online* aktivnosti, automatsko vođenje aktivnosti učenika i inteligentno stvaranje i preporučivanje materijala za učenje, mogućnost da nastavnik otkrije korisne veze između različitih načina na koje se studenti koriste informacijama, određivanje pogrješaka koje studenti istodobno čine, dobivanje detaljnih povratnih informacija o procesu e-učenja, određivanje studentskih obrazaca *online* navigacije i njihovo korištenje u evaluaciji i prilagođavanje materijala korištenih u kolegiju napretku studenata, neke su od najvažnijih učinkovitih primjena pravila pridruživanja u području obrazovanja (Garcia i sur., 2010).

Rezultati

Skup podataka korišten u ovom istraživanju dobiven je iz informacijskog sustava o studentima na Sveučilištu Celala Bayara. Skup je obuhvaćao podatke o 692 studenta dodiplomskih studija smjera Poslovna administracija na Fakultetu ekonomije i administracije. Glavni atributi tog skupa podataka bili su: klasa, dob, spol, nacionalnost, mjesto rođenja, obiteljsko i bračno stanje, posljednji stupanj obrazovanja majke, posljednji stupanj obrazovanja oca, broj braće i sestara, vrsta studija (redovni ili večernji program), primanja i status uspješnosti svakog studenta. Od ukupnog broja studenata koji su sudjelovali u istraživanju 349 ih je muškog, a 343 ženskog spola. Broj studenata koji se upisao u redoviti studijski program je 342, a ostali studenti pohađaju večernji studijski program. U ovom su se istraživanju koristila tri dobro poznata i prije spomenuta algoritma pravila pridruživanja u rudarenju podataka: Apriori, Prediktivni Apriori i Tertius algoritam. Uz njihovu pomoć izdvojena su korisna pravila pridruživanja iz skupa podataka. U provedbi eksperimentalne studije koristila se WEKA. To je softver otvorenog izvora napisan u programskom jeziku Java, a sadrži mnoštvo algoritama strojnog učenja za analizu podataka i prediktivno modeliranje (Witten i sur., 2011). Da bi se mogao primijeniti algoritam pravila pridruživanja, najprije se prethodno mora obraditi skup podataka s pomoću nenadgledanog konverzijskog filtra koji podatke pretvara iz numeričkog u nominalni oblik. Da bi se pronašla jaka pravila pridruživanja o uvjetima uspješnosti studenata, svi studenti su uglavnom podijeljeni u dvije različite skupine, ovisno o njihovim prosječnim ocjenama – u skupinu uspješnih i skupinu neuspješnih studenata. Tada su na bazu podataka primijenjena tri algoritma pravila pridruživanja. U Tablici 1 prikazana su pravila pridruživanja dobivena Apriori algoritmom. Tablica pokazuje najbolja (jaka) pravila pridruživanja dobivena algoritmom, pravila pridruživanja na osnovi dvije vrijednosti zanimljivosti: potpori i pouzdanosti. Ovdje su prikazane potpora (koja pokazuje razmjer transakcija u bazi podataka koja sadrži relevantni skup podataka) i pouzdanost (koja pokazuje vjerojatnost pronalaženja desne strane pravila u transakcijama pod uvjetom da se te transakcije nalaze i na lijevoj strani) za jaka pravila pridruživanja.

Tablica 1
Pravila pridruživanja dobivena Apriori algoritmom

Pravila	Vrijednosti potpore	Pouzdanost
Ako{Dob=19 IVrsta_programa=Redovni_program} ==>Klasa=1	73 / 73	1,0
Ako{Dob=19 IStatus_uspješnosti=Neuspješan} ==>Klasa=1	98 / 92	0,94
Ako{Spol=MuškoIPosljednji_stupanj_obrazovanja_oca=Srednja_škola} ==>Status_uspješnosti=Neuspješan	90 / 84	0,93
Ako{Dob=19} ==>Klasa=1	124 / 115	0,93
Ako{ Posljednji_stupanj_obrazovanja_majke=Srednja_škola_ili_nižiIPosljednji_ stupanj_obrazovanja_oca=Srednja_ škola_ili_nižiIVrsta_programa=Redovni_ program IStatus_uspješnosti=Neuspješan} ==>Status_Primanja=Nizak	96 / 89	0,93
Ako{Spol=MuškoIPosljednji_stupanj_obrazovanja_oca =Srednja_ škola_ili_ niži= Ivrsta_programa=Redovni_program}==>Status_primanja=Nizak	77 / 71	0,92
Ako{Dob=19 IStatus_primanja=Nizak}==>Klasa=1	75 / 69	0,92
Ako{Klasa=1 ISpol=ŽenskoIStatus_primanja =NizakIPosljednji_stupanj_ obrazovanja_oca =Srednja_ škola_ili_niži }==>Posljednji_stupanj_obrazovanja _majke= Srednja_ škola_ili_niži	88 / 80	0,91
Ako{Klasa=1 IPosljednji_stupanj_obrazovanja _majke= Srednja_ škola_ili_ nižiIPosljednji_stupanj_obrazovanja_oca= Srednja_ škola_ili_nižiIvrsta_ programa=Redovni_program}==>Status_primanja=Nizak	87 / 79	0,91
Ako{ posljednji_stupanj_obrazovanja _majke= Srednja_ škola_ili_ nižiIposljednji_stupanj_obrazovanja_oca= Srednja_ škola_ili_nižiIvrsta_ programa= Redovni_program }==> Status_primanja=Nizak	151 / 137	0,91
Ako{ posljednji_stupanj_obrazovanja_oca= Srednja_ škola_ili_nižiIvrsta_ programa= Redovni_program IStatus_uspješnosti=Neuspješan}==> Status_ primanja =Nizak	115 / 104	0,90
Ako{Dob=21 Iposljednji_stupanj_obrazovanja_oca= Srednja_ škola_ili_ niži}==>posljednji_stupanj_obrazovanja _majke= Srednja_ škola_ili_niži	90 / 81	0,90
Ako{Klasa=1 ISpol=MuškoIposljednji_stupanj_obrazovanja _majke= Srednja_ škola_ili_nižiIposljednji_stupanj_obrazovanja_oca= Srednja_ škola_ili_ niži}==>Status_primanja =Nizak	90 / 81	0,90

U Tablici 2 prikazana su pravila pridruživanja dobivena Prediktivnim Apriori algoritmom. U Prediktivnom Apriori algoritmu pravilo se dodaje ako je očekivana točnost predviđanja tog pravila među *n* najboljim pravilima. Stoga Tablica 2 uključuje i vrijednosti potpore i stopu točnosti.

Tablica 2
Pravila pridruživanja dobivena Prediktivnim Apriori algoritmom

Pravila	Vrijednosti potpore	Točnost
Ako {Dob=19 IVrsta_programa=Redovni_program} ==>Klasa=1	73 / 73	0,99488
Ako {Klasa=2 ISpol=MuškoIposljednji_stupanj_obrazovanja_oca=Srednja_škola} ==>Status_uspješnosti=Neuspješan	34 / 34	0,99411

Pravila	Vrijednosti potpore	Točnost
Ako {Spol=MuškoI posljednji_stupanj_obrazovanja_MAJKE=Srednja_ školaIposljednji_stupanj_obrazovanja_oca=Srednja_školaIVrsta_ programa=Večernji_program} ==>Status_uspješnosti=Neuspješan	26 / 26	0,99338
Ako {Klasa=2 IDob=20 Iposljednji_stupanj_obrazovanja_MAJKE= Srednja_ škola_ ili_niži Iposljednji_stupanj_obrazovanja_oca= Srednja_ škola_ili_niži IStatus_uspj ešnosti=Neuspješan}==>Status_primanja=Nizak	25 / 25	0,99324
Ako {Dob=19 IStatus_primanja=Visok} ==>Klasa=1	24 / 24	0,99309
Ako {Spol=MuškoImjesto_rođenja=IZMIR Iposljednji_stupanj_obrazovanja_ oca=Srednja_školaIStatus_primanja=Nizak} ==>Status_uspješnosti=Neuspješan	23 / 23	0,99292
Ako {Klasa=2 ISpol=MuškoIposljednji_stupanj_obrazovanja_oca= Srednja_ škola_ili_niži IVrsta_programa=Redovni_program} ==>Status_primanja=Nizak	21 / 21	0,99252
Ako {Dob=21 ISpol=MuškoIposljednji_stupanj_obrazovanja_MAJKE= Srednja_ škola_ili_niži IVrsta_programa=Redovni_program} ==>Status_primanja=Nizak	21 / 21	0,99252
Ako {Dob=21 ISpol=ŽenskoIposljednji_stupanj_obrazovanja_oca= Srednja_ škola_ili_niži IStatus_uspješnosti=Uspješan} ==>posljednji_stupanj_ obrazovanja_MAJKE= Srednja_ škola_ili_niži	21 / 21	0,99252
Ako {Dob=21 ISpol=MuškoIposljednji_stupanj_obrazovanja_oca=Srednja_škola} ==>Status_uspješnosti=Neuspješan	19 / 19	0,99201
Ako {Dob=20 ISpol=MuškoIposljednji_stupanj_obrazovanja_oca=Srednja_ školaIVrsta_programa=Večernji_program} ==>Status_uspješnosti=Neuspješan	19 / 19	0,99201
Ako {Klasa=2 Imjesto_rođenja=IZMIR Iposljednji_stupanj_obrazovanja_ oca= Srednja_ škola_ili_nižiIStatus_uspješnosti=Neuspješn} ==>Status _ primanja=Nizak	18 / 18	0,99169
Ako {Dob=21 ISpol=ŽenskoIVrsta_programa=Redovni_program IStatus_ uspješnosti=Uspješan} ==>posljednji_stupanj_obrazovanja_MAJKE= Srednja_ škola_ili_niži	18 / 18	0,99169
Ako {Spol=MuškoIposljednji_stupanj_obrazovanja_Oca=Srednja_školaIVrsta_ programa=Večernji_program} ==>Status_uspješnosti=Neuspješan	52 / 51	0,99165
Ako {Klasa=2 IDob=20 Imjesto_rođenja=IZMIR Iposljednji_stupanj_obrazovanja_ Oca= Srednja_ škola_ili_niži} ==>Status_primanja=Nizak	17 / 17	0,99134
Ako {Dob=19 ISpol=MuškoIposljednji_stupanj_obrazovanja_MAJKE= Srednja_ škola_ili_nižiIposljednji_stupanj_obrazovanja_Oca= Srednja_ škola_ili_niži} ==>Klasa=1	17 / 17	0,99134
Ako {Dob=19 Iposljednji_stupanj_obrazovanja_MAJKE=Dodiplomski} ==>Klasa=1	17 / 17	0,99092
Ako {Spol=MuškoImjesto_rođenja=AYDIN} ==>Status_uspješnosti=Neuspješan	16 / 16	0,99092
Ako {Klasa=1 IDob=21 Iposljednji_stupanj_obrazovanja_Oca= Srednja_ škola_ili_ nižiIVrsta_programa=Redovni_program} ==>posljednji_stupanj_obrazovanja_ MAJKE= Srednja_ škola_ili_niži	16 / 16	0,99092
Ako {Klasa=2 IDob=21 ISpol=ŽenskoStatus_uspješnosti=Uspješan} ==>posljednji_stupanj_obrazovanja_MAJKE= Srednja_ škola_ili_niži	16 / 16	0,99092
Ako {Dob=19 Iposljednji_stupanj_obrazovanja_MAJKE=Srednja_ školaIposljednji_stupanj_obrazovanja_Oca=Srednja_školaIStatus_ uspješnosti=Neuspješan} ==>Klasa=1	16 / 16	0,99092

Ako {Dob=19 ISpol=MuškoIposljednji_stupanj_obrazovanja_Oca=Srednja_škola} ==>Klasa=1	15 / 15	0,99043
Ako {Dob=23 Iposljednji_stupanj_obrazovanja_Oca= Srednja_ škola_ili_ nižiIStatus_primanja=Nizak} ==>posljednji_stupanj_obrazovanja_MAJKE= Srednja_ škola_ili_niži	15 / 15	0,99043
Ako {Klasa=2 IDob=20 ISpol=MuškoIposljednji_stupanj_obrazovanja_Oca= Srednja_ škola_ili_niži} ==>Status_primanja=Nizak	15 / 15	0,99043
Ako {Dob=21 ISpol=MuškoIposljednji_stupanj_obrazovanja_Oca= Srednja_ škola_ili_nižiIVrsta_programa=Redovni_program} ==>Status_primanja=Nizak	15 / 15	0,99043

Tablica 3 prikazuje pravila pridruživanja dobivena Tertius algoritmom.

Tablica 3

Pravila pridruživanja dobivena Tertius algoritmom

Pravila
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = ERZURUM ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = ELAZIG ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = BITLIS ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = ANTALYA ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = KARS ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = MALATYA ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = KAHRAMANMARAS ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = VAN ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = ERZINCAN ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>DOB = 28 ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = ARTVIN ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = AFYON ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = EDIRNE ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = HATAY ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži

Pravila
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = HATAY ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = MUGLA ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>DOB = 27 ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = CANAKKALE ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>posljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_nižiILIKLASA = 3
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>posljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = KUTAHYA ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = ANKARA ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = DIYARBAKIR ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = SIVAS ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = KASTAMONU ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = SAKARYA ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži
Ako { posljednji_stupanj_obrazovanja_majke = Srednja_škola_ili_nižiIStatus_primanja = Nizak } ==>mjesto_rođenja = YOZGAT ILIposljednji_stupanj_obrazovanja_oca = Srednja_škola_ili_niži

Rasprava

Za Apriori algoritam odabrana su pravila sa stopom pouzdanosti višom od 90 %. Za Prediktivni Apriori algoritam odabrana su pravila sa stopom točnosti višom od 99 %, a za Tertius algoritam odabrana su pravila sa stopom potvrde višom od 79 %.

U Tablici 1 prikazano je 13 pravila dobivenih s pomoću Apriori algoritma. Kako se može vidjeti iz rezultata, najbolje utvrđeno pravilo s najvišom vrijednošću pouzdanosti upućuje na to da su studenti u dobi od 19 godina koji su se upisali u redoviti studijski program na svojoj prvoj godini dodiplomskog studija. Slično tome, za drugo i četvrto pravilo dob ima vrijednost 19 i klasu 1 u svojem prethodniku i sljedbeniku pojedinačno. Treće pravilo navedeno u Tablici 1 pokazatelj je pridruživanja spola studenata, posljednjeg stupnja obrazovanja koji su završili njihovi roditelji i njihova statusa uspješnosti. Prema tom pravilu može se s visokom stopom pouzdanosti očekivati da će studenti muškog spola s očevima koji imaju srednju stručnu spremu biti neuspješni u kolegijima. Peto pravilo pridruživanja pokazuje da se očekuje da neuspješni studenti koji pohađaju redovite studijske programe i čiji roditelji imaju relativno nizak stupanj obrazovanja imaju niska obiteljska primanja. Slično tome, još pet pravila sadrži nizak status primanja u svojem nastavku. Osmo pravilo navodi da se očekuje da će studentice prve godine studija koje imaju niska primanja i očeve sa srednjom stručnom spremom ili niskim stupnjem obrazovanja imati majke koje su također završile srednju školu ili niži stupanj obrazovanja. Slično tom pravilu, dvanaesto pravilo također naglašava vezu sa stupnjem obrazovanja roditelja.

Prediktivnim Apriori algoritmom stvoreno je 25 pravila pridruživanja. Budući da se Apriori i Prediktivni Apriori algoritmi temelje na različitoj metrici, pravila stvorena Prediktivnim Apriori algoritmima nisu potpuno jednaka onima koja su dobivena Apriori algoritmom. Kada se pravila pridruživanja grupiraju prema svojim sljedbenicima, može se primijetiti da pravila stvorena Prediktivnim Apriori algoritmom otkrivaju rezultate povezane s četiri uvjeta: klasom, statusom uspješnosti, statusom primanja i posljednjim stupnjem obrazovanja majke. Pravila koja upućuju na činjenicu da su studenti na prvoj godini studija, imaju dobnu vrijednost 19 u svojim prethodnicima. Kao što je i dobiveno Apriori algoritmom, pravila koja se tiču statusa uspješnosti studenata uvelike upućuju na to da su studenti čiji očevi imaju srednju stručnu spremu općenito neuspješni. Štoviše, pravila koja sadrže status primanja u svojim sljedbenicima obično u svojim prethodnicima pokazuju nizak stupanj obrazovanja roditelja. Posljednji važan podatak koji je dobiven a pomoću Prediktivnog Apriori algoritma pokazuje da oni studenti čiji očevi imaju niži stupanj obrazovanja vjerojatno imaju i majke čiji je stupanj obrazovanja također niži. Kao što se može primijetiti u pravilima navedenima u Tablici 3, pravila stvorena Tertius algoritmom nisu tako čvrsta kao ona dobivena Apriori ili Prediktivnim Apriori algoritmom, jer ta pravila općenito u svojim sljedbenicima imaju veznike. Sva pravila Tertius algoritma upućuju na to da studenti s niskim obiteljskim primanjima i s majkama koje imaju niži stupanj obrazovanja imaju i očeve koji također imaju niži stupanj obrazovanja.

Glavni je cilj eksperimentalne studije izdvojiti čvrsta pravila pridruživanja u vezi sa statusom uspješnosti studenata. Međutim, studija je ograničena upotrebom demografskih atributa u formuliranju modela. Iako demografski atributi mogu doprinijeti točnijem određivanju statusa uspješnosti studenata, oni sami po sebi nisu dovoljni. Da bi se izradio stabilniji model, u obzir bi trebalo uzeti i atribute kao što su pojedinačne ocjene koje studenti imaju iz različitih kolegija i njihove ocjene na prijemnom ispitu. Međutim, informacijski sustav o studentima iz kojega su se izdvajali skupovi podataka nije baš fleksibilan u pružanju takvih atributa. Mnogi potencijalno korisni atributi ili ne postoje ili im nedostaju određene vrijednosti. Stoga se u ovom istraživanju nisu mogla dobiti neka zanimljiva pravila pridruživanja do kojih bi se inače moglo doći.

Sve u svemu, stvorena su mnogobrojna i različita pravila pridruživanja s pomoću prije spomenutih algoritama. Navedeni rezultati dobiveni Apriori i Prediktivnim Apriori algoritmima upućuju na to da roditelji neuspješnih studenata imaju niska primanja ili je njihov stupanj obrazovanja relativno nizak (srednja stručna sprema ili raniji stupanj obrazovanja). Stoga, općenito govoreći, što su njihovi roditelji obrazovaniji, to su studenti uspješniji u sveučilišnom obrazovanju. Osim toga, pravila također pokazuju povezanost obiteljskih primanja i stupnja obrazovanja. Još jednom ćemo reći, u obiteljima s niskim primanjima roditelji obično imaju relativno nizak stupanj obrazovanja, pa su studenti iz takvih obitelji uglavnom neuspješni na studiju. Strateški menadžment je disciplina koja se bavi odabirom odgovarajućih strategija, planova, primjena i aktivnosti da bi se organizacijama omogućilo postizanje ciljeva. Rudarenje podataka je uspješan alat za prikupljanje učinkovitih rješenja za strateški menadžment. Područje obrazovanja obiluje raznovrsnim podacima, pa stoga rudarenje podataka može dati učinkovita rješenja za strateški menadžment u području obrazovanja.

Zaključak

Rudarenje podataka u obrazovanju ima za cilj poboljšati kvalitetu procesa učenja. Među mnogim mogućim područjima primjene rudarenja podataka u obrazovanju, modeliranje studenata ima glavnu ulogu u ostvarivanju željenih ciljeva za poboljšanje kvalitete obrazovanja. Potreba i pogodnost metoda rudarenja podataka u strateškom menadžmentu u poslovnim organizacijama dobro je poznato pitanje. U ovom su radu istraživači ispitali upotrebu algoritama pravila pridruživanja u rudarenju podataka u obrazovanju koristeći se podacima o studentima koji su prikupljeni iz informacijskog sustava o studentima na Sveučilištu Celala Bayara. Skup podataka uglavnom je karakteriziran demografskim atributima studenata. Iako demografski podaci sami po sebi možda ne bi bili dovoljni da se potpuno modelira status uspješnosti studenata, s pomoću algoritama su ipak pronađena neka zanimljiva pravila pridruživanja. Ta dobivena pravila pridruživanja ukazuju na to da postoji jaka veza između stupnja obrazovanja roditelja, obiteljskih primanja i akademskog uspjeha. Prema tim rezultatima, pouzdano se može očekivati da će studenti s niskim obiteljskim primanjima ili studenti čiji roditelji imaju nizak stupanj obrazovanja vjerojatno biti neuspješni u akademskom obrazovanju. Stoga bi se već od prve godine dodiplomskog studija trebala organizirati savjetovanja za studente za koje se očekuje da će biti neuspješni na studiju. Kako je već prije spomenuto, ovu bi studiju u budućnosti trebalo obogatiti poboljšanim glavnim atributima skupa podataka.