

Looking Relationships between Pre-admission Factors and Students' Performance Using Association Rules

Silvia Beatriz González-Brambila, Josué Figueroa-González

Universidad Autónoma Metropolitana,
Azcapotzalco, Mexico

{sgb, jfgo}@correo.azc.uam.mx

Abstract. High school institutions are very demanded, thousands of students desire entering to them and universities are looking for the best candidates considering several criteria like admission tests, former level average or extra curricular activities. Educational Data Mining is a branch of Data Mining which processing a lot of educative data, allows finding interesting patterns which allow taking decisions for solving or improving educative aspects. This work presents the use of Data Mining Association Rules technique for finding rules related with students pre-admission characteristics and their first scholar year performance. Goal of the paper is analyzing if using as single criteria the highest mark in an admission test allows the Mexican Autonomous Metropolitan University choosing students which will have the best performance. Results show that even if admission test mark is relevant, there exist other factors, specially the former level average, which should be considered in the admission process.

Keywords. Association rules, educational data mining, enrollment criteria, student performance, university admission.

1 Introduction

Universities use different enrollment criteria for choosing students that will enter to a superior education. Entrance examination, personal interviews, scholar performance on the previous levels, extracurricular activities, etc, are some of the criteria considered for accepting a student. The admission process looks up for choosing the best students, so they can finish their studies with an acceptable performance; however it's not clear if these criteria allow universities to make it.

With the increase in the use of technology, it's possible to store a lot of data from students, which being processed can lead to interesting information that can be helpful for making decisions. Data Mining (DM) is the process of analyzing large amount of data for discovering patterns that lead to knowledge [6], applying this process to academic data is known as Educational Data Mining (EDM) [2].

DM offers a lot of algorithms according a desired goal; in particular association techniques allow discovering relationships among several data. Association rules show the relationship between certain event or events called antecedents, and another event, or events, called consequence.

In this work is applied the Association Rules technique for discovering relationships between some entrance criteria and the performance of engineering students over their first year at the Autonomous Metropolitan University (UAM). This one, uses as enrollment criteria a single admission exam where students with best marks are considered for entering university. This mark must be higher or at least, very near, to 600 points (from a maximum of 1000), other prerequisite is that medium high level average, in Mexico is the former level to superior education, must be at least 7.0 (from a maximum of 10). Consider that not all the students which fulfill these criteria are accepted, only a small number of candidates, for this reason it's important to know if a single admission test is the best way for choosing from all the students that look up for entering UAM.

At UAM, first year is called the General Branch Level (GBL) and is composed by 3 quarters (scholar period at UAM) where all engineering students study the same topics and it's the period with the highest level of desertion and the one which take students most time for finishing it. For these reasons, it's performed a study than can find relationships between some criteria besides the admission test and the performance of students.

Goal of this work is analyzing the rules obtained from the relationship between the single entrance criteria, admission test, and other which now are not considered like medium high school average, student age and school of origin with the academic performance of students during their first scholar year. Performance is measured using Grade Point Average (GPA) and the amount of approved credits during this period. GPA is a measure obtained from the relationship of approved topics multiplied and the obtained mark against attempted credits and it's obtained using Eq. (1):

$$GPA = \frac{\sum(credits * mark)}{attempted\ credits}, \quad (1)$$

where:

- credits is the amount of credits of a subject,
- mark is the obtained mark in a topic (only for approved ones),
- attempted credits is the total of credits coursed, approved or not, in a certain period.

This will allow determining if there is a relationship between student performance and entrance exam, or if it's necessary to consider other factors that actually are being ignored in the process of choosing students at UAM.

Paper is structured as follows: Section 2 presents related works about the topic. Section 3 presents the processes of obtaining data and generating the association rules. Section 4 shows the analyzis of the obtained rules for both performances, considering GPA and amount of approved credits. Section 5 presents Conclusions and Future works.

2 Related Work

The main topic studied with EDM is student performance over a certain period or for all studies, related with student performance over first year is [4], where linear regression and Decisions Trees are used for predicting one of three categories created according students first year performance over a set of the French Community of Belgium. Results show that academic factors like average in last year of secondary school, time of studying mathematics and class attendance have a strong relationship with the classification of a student in one of the three categories based in its performance.

Besides this topic, several works related with the analyzis of enrollment criteria over students' performance have been performed. This is an interesting topic for universities because they are always looking for the best way of choosing their new students.

In [8], the relationships between several factors including age, GPA of previous studies, admission test marks and performance over first semesters in a former level with the first semester in a nursing school program are obtained. Used technique is regression analysis, results shown that admission test mark is the criteria which has most relevance with student performance.

In [9], was performed a study about the influence of academic and personal factors over first semester and first year academic performance in an Ethiopian university. Used techniques were correlation analysis and linear regression. Results shown that former grade and admission test have the greatest influence over student performance, also socio economic variables has certain relationship with student achievement.

Finally, in [3], the importance of general tests, called State Matura exams, and the average in former level have over the performance of student in two faculties in Croatia is studied. Matura exams include Croatian language, mathematics and a foreign language. Used techniques are also multiple linear regression and correlation analysis. Obtained results are not considered statistical relevant and it's concluded that Matura exams are not useful for predicting students performance.

As can be seen, all of the analyzed works used the correlation analysis and multiple regressions for predicting the performance of students. In this work, association rules are used, not for predicting, but for identifying relationships among several factors with student performance.

3 Applying Data Mining for Discovering Rules

Considering the steps of several DM methodologies [1], once the goal is defined, the process for obtaining knowledge must include at least the following steps: obtaining and preparing data, applying DM algorithms, in this case Association Rules algorithm, and finally, analyzing results for obtaining knowledge.

3.1 Obtaining and Preparing Data

UAM stores student data into two sources, the General File of Students (AGA for its acronym in Spanish) which contains a lot of personal and academic data from all the students in the university. Second source is the academic record, called kardex in UAM, these data correspond specifically to the academic history of every student, each register is composed by the registration number of a student, id of a topic, obtained mark and the period when the student studied the topic. For this work, were considered Computer Engineering students from years 2010 to 2016 for a total of 538 students. Students from previous years were not considered because their marks, in 2010, a major modification in the students' plan was performed, so marks in several subjects were change to EQ (Equivalent) and the original mark wasn't obtained.

From the AGA, were obtained the following data related with pre-admission characteristics:

- School of Origin (SCH). Indicates the medium high level school from where a student proceeds.
- Medium high school average (AVG). Represents the average obtained by a student in the medium high level.
- Entrance age (AGE). Is the age of the student at the moment of being accepted into the university.
- Admission exam mark (PEX). Total of points earned by a student in the admission exam that UAM applies for candidates of their engineering programs.

From kardex, were obtained the studied topics and obtained marks from the first year, in UAM, the scholar calendar is divided in quarters, three quarters per year, so, were considered the topics of the first three quarters according the entrance quarter (university has two periods for entrance). Marks at UAM are assigned with a letter, these are presented in Table 1; letters are presented according their acronym in Spanish. While Not Approved mark stands for 5, at the moment of obtaining GPA, the used value was 0.

Table 1. Marks and their numeric equivalent at UAM.

Mark	Numeric Equivalent
MB (Very Good)	10
B (Good)	8
S (Sufficient)	6
NA (Not Approved)	5

Data preparation, or data cleaning, involves obtaining GPA and the amount of credits approved for each student during its first year. For obtaining the amount of credits, were added the credits of each approved topic during the first

year (three quarters) of every student. Credits of each topic were taken from the study plan of Computing Engineering [7].

For obtaining better results using the Association Results technique, data were categorized, so, were created categories for admission test, medium high school average, student age, school of origin, GPA and amount of approved credits. Next, are presented the used categories, in each one, between parentheses, values that appears in the rules.

For creating admission test categories were considered the range of values presented in Table 2, consider that the highest value was 909 (from a maximum of 1000) and the lowest, 548 points.

Table 2. Categories of admission test obtained mark.

Range of Values	Equivalent
More or equal than 880	High (EX_HIGH)
From 770 to 880	Medium (EX_MED)
Less than 770	Low (EX_LOW)

Categories used for medium high school average are presented in Table 3. These categories were created considering the most common marks scale used at UAM. Here the values run from 7.0 to 10, remembering that 7.0 is the minimum value for accessing the university.

Table 3. Categories of medium high school average.

Range of Values	Equivalent
More or equal than 8.8	High (AVG_HIGH)
From 7.7 to 8.8	Medium (AVG_MED)
Less than 7.7	Low (AVG_LOW)

Categories for entrance age are presented in Table 4. Ages from accepted students were from 16 to 41 years old. The most common entrance age in Mexico to superior level runs from 18 to 20 years old.

Table 4. Categories of entrance age.

Range of Values	Equivalent
Less than 18	Group 1 (GPR_1)
From 18 to 20	Group 2 (GPR_2)
From 20 to 25	Group 3 (GPR_3)
Greater than 25	Group 4 (GPR_4)

Categories of GPA are presented in Table 5. Obtained GPA of processed students runs from 0.0 (meaning that a student didn't approve any topic) to 9.9.

Table 5. Categories from Grade Point Average.

Range of Values	Equivalent
More or equal than 8.8	High (GPA_HIGH)
From 7.6 to 8.7	Medium (GPA_MED)
From 6.0 to 7.5	Low (GPA_LOW)
From 0.0 to 5.9	Very Low (GPA_VLOW)

Finally, categories for approved credits are presented in Table 6. The expected amount of credits considering the Computing Engineering study plan for the first year is 105; however, at UAM, students can study their topics in a different order than the presented in the study plan, they even can approve more credits; for this reason, were considered the estimated amount of credits expected for quarter for the classification: 1st quarter 32 credits, 2nd quarter 33 and 3rd quarter, 40. Credits for analyzed students run from 0 to 120 credits.

Table 6. Categories from approved credits.

Range of Values	Equivalent
From 80 to more than 105	High (CRE_HIGH)
From 32 to less than 80	Medium (CRE_MED)
Less than 32	Low (CRE_LOW)

This classification means that a student with a low performance, approved less than 32 credits in its first year (less than the credits of 1st quarter), a medium performance represents approving from 32 to 80 credits (the sum of credits of first, second quarter and some credits of third quarter) during the first year. Finally, the high performance is from 80 to more than 105, that represents approving all, or almost, the topics from first, second and third quarter during the first year.

Also, was considered the school of origin of every student, schools were grouped according the main institution system they belong. However, for privacy reasons, in this work are presented with generic names, School A (SCH_A) to School I (SCH_I).

3.2 Obtaining Association Rules

For generating Association Rules, was used the *Apriori* algorithm; with this algorithm, many rules are generated, so it's necessary to choose the most representative, there are several methods for choosing the best rules [5], the most

used value for this is lift, which represents the occurrence frequency of a set of elements X and Y respecting an expected value and is defined in Eq. (2):

$$lift(X \Rightarrow Y) = \frac{support(X \Rightarrow Y)}{support(X) * support(Y)}, \tag{2}$$

where support is defined as the percentage of transactions that contains X in a set of transactions D.

For evaluating the importance of a rule, are considered the following criteria: if value of lift is equal to 1, means that the relationship could be due to a random consequence. If lift is greater than 1, represents that exists a strong relationship among antecedents and consequences (X and Y are called complements). Finally, if lift is less than 1, there is not a relationship (X and Y are called substitutes).

Were obtained rules considering as antecedents different combinations of values for admission test, medium high school average, student age and school of origin, consequence events were GPA and amount of approved credits separately. Rules for GPA are presented in Tables 7 to 9 and the ones for total of credits are shown in Tables 10 to 12.

Table 7. Rules obtained for high performance measured with GPA.

Rule	Support	Confidence	Lift
{EX_HIGH,GPR_2,SCH_I} ⇒ {GPA_HIGH}	0.001879	1.0	13.641
{AVG_MED,EX_HIGH,GPR_2} ⇒ {GPA_HIGH}	0.001879	1.0	13.641
{AVG_HIGH,SCH_D,EX_LOW} ⇒ {GPA_HIGH}	0.001879	1.0	13.641
{AVG_HIGH,EX_LOW,GPR_3,SCH_E} ⇒ {GPA_HIGH}	0.001879	1.0	13.641
{AVG_HIGH,SCH_D,GPR_2} ⇒ {GPA_HIGH}	0.005639	0.75	10.230

4 Analyzing Association Rules

Next are presented the analysis of the generated rules for each case.

4.1 Interpreting Rules Related with Performance Measured Using GPA

Analyzing the antecedents that tends to a high performance considering GPA, rules from Table 7, it's clear that a high mark in the admission exam and a high average in former level (EX_HIGH) and AVG_HIGH) are associated with a higher performance, this can be a rule called trivial, however, it's interesting that in two rules exist a low mark in the admission exam (EX_LOW) that appears with a high average. Talking about entrance age, two groups can be found in the rules, the group of the most common entrance age (GPR_2) and another group which age is little older (GPR_3).

Table 8. Rules obtained for medium performance measured with GPA.

Rule	Support	Confidence	Lift
{EX_HIGH,SCH_G} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{SCH_D,EX_MED,GPR_4} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{AVG_MED,SCH_D,EX_MED} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{AVG_MED,EX_MED,SCH_H} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{EX_MED,GPR_3,SCH_H} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{SCH_C,EX_MED,GPR_2} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{AVG_MED,SCH_C,EX_MED} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{AVG_HIGH,SCH_C,GPR_4} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{EX_MED,GPR_3,SCH_I} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{AVG_LOW,SCH_A,EX_MED} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{AVG_LOW,SCH_A,GPR_4} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{AVG_MED,EX_MED,GPR_4} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{AVG_LOW,EX_MED,SCH_F} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{AVG_LOW,SCH_C,EX_LOW,GPR_3} ⇒ {GPA_MED}	0.001879	1.0	3.5231
{SCH_A,EX_MED,GPR_3} ⇒ {GPA_MED}	0.005639	0.75	2.6423

Table 9. Rules obtained for low performance measured with GPA.

Rule	Support	Confidence	Lift
{GPR_4,SCH_H} ⇒ {GPA_LOW}	0.001879	1.0	3.4322
{AVG_LOW,GPR_4,SCH_I} ⇒ {GPA_LOW}	0.003759	1.0	3.4322
{EX_MED,SCH_F,GPR_4} ⇒ {GPA_LOW}	0.001879	1.0	3.4322
{AVG_LOW,EX_MED,GPR_2} ⇒ {GPA_LOW}	0.001879	1.0	3.4322
{AVG_MED,EX_MED,GPR_2,SCH_G} ⇒ {GPA_LOW}	0.001879	1.0	3.4322
{EX_LOW,GPR_4,SCH_I} ⇒ {GPA_LOW}	0.0075187	0.8	2.7458
{AVG_LOW,GPR_3,SCH_I} ⇒ {GPA_LOW}	0.0075187	0.8	2.7458

After reviewing rules in Table 8, the ones related with a medium performance considering GPA, they show something interesting, the average of the former level it's medium (AVG_MED), there is a set of rules containing high former level average (AVG_HIGH), but appears with the oldest group of age (GPR_4) which doesn't appear in the high GPA performance. Also, in this set of rules, doesn't appear neither high (EX_HIGH) not low (EX_LOW) marks in the admission test.

Analyzing the seven rules in Table 9, associated with a low performance, can be seen that five of them contain one of the groups (GPR_3 and GPR_4) which are older than the common entrance age. Also the former level average is classified in low or medium (AVG_MED and AVG_LOW). About admission exam, only were found medium marks (EX_MED).

From this analysis, can be told that the average in the former level, in this case, the medium high level, can lead to choose students which trend to have a better performance. The fact that only one high or low admission exam marks and none high or low former average appears in the rule of medium performance, strengths the theory that a high performance can be associated with a high mark

Table 10. Rules obtained for high performance measured with amount of credits.

Rule	Support	Confidence	Lift
{EX_HIGH,GPR_2} ⇒ {CRE_HIGH}	0.003759	1.0	2.67336
{EX_MED,SCH_I} ⇒ {CRE_HIGH}	0.01357	1.0	2.67336
{AVG_HIGH,SCH_C,GPR_4} ⇒ {CRE_HIGH}	0.001879	1.0	2.67336
{AVG_HIGH,GPR_2,SCH_E} ⇒ {CRE_HIGH}	0.001879	1.0	2.67336
{AVG_HIGH,EX_LOW,SCH_E} ⇒ {CRE_HIGH}	0.003759	1.0	2.67336
{EX_MED,GPR_2} ⇒ {CRE_HIGH}	0.03571	0.8260	2.20843
{AVG_HIGH,SCH_D} ⇒ {CRE_HIGH}	0.007518	0.8	2.13869
{AVG_HIGH,EX_MED} ⇒ {CRE_HIGH}	0.04699	0.7575	2.02527
{AVG_LOW,GPR_2,SCH_I} ⇒ {CRE_HIGH}	0.005639	0.75	2.00502
{AVG_HIGH,SCH_A,GPR_4} ⇒ {CRE_HIGH}	0.005639	0.75	2.00502
{EX_MED,SCH_E} ⇒ {CRE_HIGH}	0.0206	0.7333	1.96046
{AVG_HIGH,SCH_I} ⇒ {CRE_HIGH}	0.0244	0.7222	1.93076

in the admission test and a high average in the medium high level.

As can be seen, the entrance age also plays an important role in the performance of the student, showing that groups with an older age, specially Group 4, tends to have a lower performance.

4.2 Interpreting Rules Related with Performance Considering the Mmount of Approved Credits

Can be considered that as highest is the GPA, highest is the amount of credits, however there it's not relationship, for example: a GPA of 10 (the maximum value) can be obtained approving one topic with the highest mark (MB), and a GPA of 6 can be obtained approving more than one topic with the minimum approving mark (S). Also, it's supposed that the characteristics associated with a high performance considering GPA, should be the same for a high performance considering the amount of credits, but to assure this, the set of rules related with the amount of credits are also analyzed.

As rules associated with high amount of approved credits, Table 10, show, having a high average in the former level (AVG_HIGH) appears in most of the rules with the highest value of lift. Here the relationship of the mark in the admission exam it's not as clear as with the performance considering GPA. Were obtained rules with each one of the possible values for the exam (EX_HIGH, EX_MED and EX_LOW). Bigger amount of credits it's associated also with several entrance age groups, but it's clear that the common entrance age group (GPR_2) is present in most of the rules, even with an entrance exam mark or former level average which are not high, while another group (GPR_4) is present only with high average in medium high level.

Looking the rules of Table 11, the ones related with medium performance, it's clear that former level average has an important role in the amount of accumulated credits, most of the rules contain low or medium average (AVG_LOW and AVG_MED). Also the mark in the admission exam has many low and medium

Table 11. Rules obtained for medium performance measured with amount of credits.

Rule	Support	Confidence	Lift
{EX_HIGH,GPR_3} ⇒ {CRE_MED}	0.003759	1.0	2.1111
{SCH_D,GPR_3} ⇒ {CRE_MED}	0.007518	1.0	2.1111
{SCH_D,EX_MED,GPR_4} ⇒ {CRE_MED}	0.001879	1.0	2.1111
{AVG_LOW,GPR_3,SCH_I} ⇒ {CRE_MED}	0.009398	1.0	2.1111
{AVG_LOW,SCH_A,GPR_4} ⇒ {CRE_MED}	0.001879	1.0	2.1111
{EX_MED,GPR_4,SCH_F} ⇒ {CRE_MED}	0.001879	1.0	2.1111
{AVG_LOW,EX_MED,GPR_4} ⇒ {CRE_MED}	0.001879	1.0	2.1111
{AVG_MED,EX_MED,GPR_4} ⇒ {CRE_MED}	0.001879	1.0	2.1111
{AVG_LOW,EX_MED,SCH_G} ⇒ {CRE_MED}	0.003759	1.0	2.1111
{AVG_LOW,GPR_4,SCH_F} ⇒ {CRE_MED}	0.001879	1.0	2.1111
{AVG_LOW,SCH_C,EX_LOW,GPR_3} ⇒ {CRE_MED}	0.001879	1.0	2.1111
{AVG_MED,EX_MED,GPR_2,SCH_G} ⇒ {CRE_MED}	0.001879	1.0	2.1111
{AVG_MED,SCH_D} ⇒ {CRE_MED}	0.011278	0.85714	1.8095
{AVG_LOW,GPR_2,SCH_F} ⇒ {CRE_MED}	0.007518	0.8	1.6888
{EX_LOW,GPR_3,SCH_H} ⇒ {CRE_MED}	0.005639	0.75	1.58333
{AVG_MED,SCH_B,GPR_3} ⇒ {CRE_MED}	0.005639	0.75	1.5833
{AVG_HIGH,SCH_A,EX_LOW,GPR_3} ⇒ {CRE_MED}	0.005639	0.75	1.5833
{SCH_D,EX_LOW} ⇒ {CRE_MED}	0.009398	0.71428	1.5079
{AVG_MED,EX_LOW,SCH_H} ⇒ {CRE_MED}	0.009398	0.71428	1.5079
{AVG_MED,GPR_2,SCH_I} ⇒ {CRE_MED}	0.009398	0.71428	1.5079

Table 12. Rules obtained for low performance measured with amount of credits.

Rule	Support	Confidence	Lift
{SCH_B,GPR_4} ⇒ {CRE_LOW}	0.001879	1.0	6.5679
{AVG_LOW,SCH_C,EX_MED} ⇒ {CRE_LOW}	0.001879	1.0	6.5679
{SCH_C,EX_MED,GPR_3} ⇒ {CRE_LOW}	0.001879	1.0	6.5679
{AVG_LOW,GPR_4,SCH_G} ⇒ {CRE_LOW}	0.001879	1.0	6.5679
{AVG_HIGH,EX_LOW,GPR_2,SCH_H} ⇒ {CRE_LOW}	0.001879	1.0	6.5679

values (EX_LOW and EX_MED). About ages, it's not clear as the rules obtained for performance considering GPA, but also can be seen that older groups (GPR_3 and GPR_4) tends to have a lower performance than the most common entrance age (GPR_2) which it's related with the biggest amount of accumulated credits.

Reviewing the rules related with low performance, Table 12, as in the performance measured using GPA, also appears the groups of older ages (GPR_3 and GPR_4). More rules contains the lower average in the former level (AVG_LOW) and not a high mark in the admission exam, appearing only medium and low marks (EX_MED and EX_LOW).

5 Conclusions

Goal of this work was to analyze if the way that UAM is choosing their students, only the mark in the admission exam, allows choosing those ones which will have

a good performance. It's evident that a high mark in the admission exam is important, however, it's also clear that for a high performance, considering both GPA and accumulated credits can be useful to consider the former level average, not only as a prerequisite, but as a criteria that can be included in some way, maybe as a percentage, in the admission process.

It's also clear that entrance age plays an important factor over the performance of students, groups of ages greater than the common entrance age tends to have a lower performance, specially the Group 4 (greater than 25 years old). This has a good reason, at that age, it's common that a student has a job and maybe a family, so the time for investing in its studies is less than in a young person which possibly has the support of a relative. UAM gives opportunity to all people for trying to enter, so this criteria cannot be considered as a way for choosing students.

Because of confidentiality reasons, names of origin schools were not included, however, can be mentioned that they don't have a relevant impact over the performance. Besides, as institution, UAM cannot reject a student due to the school it comes, for this reason, this criteria, as the entrance age, can't be considered for choosing students.

Future works consider applying this technique not only for the first year, but for all the time students take for finishing their career. This involves including more variables, because in four or five years exist a lot of situations that can occur, like having a family, a sickness or getting a job while it's less common that this situations happen in the first year. Also, a similar study classifying students according their performance considering factors like average or time for finishing their studies can be interesting for university. This process can be applied not only for analyzing Computer Engineering students, but for all Engineering at UAM, other study areas like Social or Biological Sciences and Art and Design Studies can be of interest for UAM authorities. Also, this work can be applied to even other universities.

Acknowledgments. Thanks to Academic Secretary and Scholar System of UAM for providing the data used in this work.

References

1. Azevedo, A. I. R., Santos M. F.: KDD, semma and CRISP-DM: A parallel overview. *IADS-DM*, 182–185 (2008)
2. 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)* 6, 601–608 (2010)
3. MunjarD., Keček D., Matotek D.: Relationship between enrolment criteria and first-year students' study-success. In: *3rd Human and Social Sciences at the Common Conference-HASSACC*, pp. 111–115 (2015)
4. Superby J. F., Vandamme J. P., Meskens N.: Determination of factors influencing the achievement of the first-year university students using data mining methods. *Workshop on Educational Data Mining* (2006)

5. Sheikh L. M., Tanveer B., Hamdani M. A.: Interesting measures for mining association rules. In: Proceedings of INMIC 2004 8th International, pp. 641–644 (2004)
6. Maimon O., Rokack L.: Data Mining and Knowledge Discovery Handbook. Springer Science+Business Media (2010)
7. Universidad Autónoma Metropolitana, División de Ciencias Básicas e Ingeniería, http://cbi.azc.uam.mx/es/CBI/Planes_Programa_Estudio_Com
8. Kowitlawakul Y., Brenkus R., Dugan N.: Predictors for success for first semester, second-degree Bachelor of Science in Nursing students. International journal of nursing practice 19, 38–43 (2013)
9. Zebdewos Z., Nasser A., Fisseha M.: Predictors of academic achievement for first-year students. The case of Wolaita-Soddo University Ethiopia study-success enrolment. European Scientific Journal 28, 160–173 (2015)