Academia Journal of Educational Research 6(12): 279-289, December 2018
DOI: 10.15413/ajer.2018.0717
ISSN 2315-7704

Research Paper

Academic achievement profiles: An intelligent predictive model based on data mining

Accepted 20th December, 2018

ABSTRACT

It is well known that academic achievement is one of the key aspects in the development of educational activities and it strongly determines the chances of success during and after a university career. It is therefore important to try and effectively monitor students' performance in order to prevent problems from emerging, as well as, to be able to provide academic coaching when the performance is not adequate. The aforementioned problem-anticipation possibility is closely related to the ability to predict the most probable situation based on concrete information. In an academic achievement framework, it is desirable to be able to predict students' performance considering concrete individual parameters. This work outlines the results obtained by an academic-achievement prediction model based on data mining algorithms which uses socio-economic information as well as, students' grades. The tests were carried out at National Technological University, Resistencia Regional Faculty (UTN-FRRe), during the AED-Algoritmos y Estructuras de Datos (Algorithms and Data Structures) class throughout the 2013, 2014, 2015 and 2016 terms. The results obtained confirmed adequate behaviour of the model which has been validated for both description and prediction of academic achievement profiles.

Key words: academic achievement, student profiles, data mining, machine learning.

D. L. La Red Martinez*, M. Karanik and M. E. Giovannini

Resistencia Regional Faculty, National Technological University, French 414, (3500) Resistencia, Chaco, Argentina.

*Corresponding author. E-mail: isi.comunicaciones@frre.utn.edu.ar. Tel: +54-9379-4638194.

INTRODUCTION

Currently, on grounds of exponential growth of loads of information, there exists a strong tendency to obtain knowledge from available data. This tendency arises in most organisational, entrepreneurial, governmental and educational environments. Works which use computer technologies have been carried out in the educational field to analyse data associated to the running and administration of institutions (De Moortel and Crispeels, 2018; Dika and Hamiti, 2011; Jan and Contreras, 2011), as well as, to improve educators and students' performance and academic capability (Cerretani et al., 2016; Gökalp, 2010; Hamidi and Jahanshaheefard, 2018; Kerimbaeva et al., 2014; Spiegel and Rodríguez, 2016).

Students' academic capability clearly conditions their performance and as a consequence, their chances of proper development of their careers since students' productivity are closely related to their performance (Maletic et al., 2002). This academic performance is conditioned by the acquisition of knowledge which is necessary to overcome the different stages of their academic training by extra institutional activities and their own individual, social and behavioural abilities.

Even if academic performance is shown in grades obtained in tests, collecting information to detect and correct cognitive issues is not an easy task. It is necessary to analyze socio-cultural and economic factors as well as, educational performance background to establish students' performance profiles that enable decision making during the process (Tinto, 1993).

The relevance of identifying profiles is based on the fact that studying prospects are established by students' performance along the initial years of their university careers. Hence, it is highly desirable that universities focus on applying strategies to keep those students who present some kind of academic impairment (Oloriz et al., 2007).

Several works related to evaluation of academic performance use the grades obtained during admission to university and the data analyzed with statistical techniques (Jiménez et al., 2000). Other alternatives to determine relevant aspects of academic performance are analysis of performance indicators based on individual data (Molina et al., 2004), and even the use of the production function approach (Di Gresia, 2007).

There are also several studies on regression in academic success/failure prediction that use objective characteristics, such as class attendance and subjective characteristics, such as class involvement (García and San, 2001). Another approach based on Data Warehouse and Data Mining techniques (Martínez et al., 2017) has shown interesting results that explain the student's development on grounds of their performance along the first year.

Martínez et al. (2017) examines academic achievement in the AED-Algoritmos y Estructuras de Datos (Algorithms and Data Structures) class. This is a key subject in the System Engineering (ISI - Ingeniería en Sistemas de Información) career at National Technological University, Resistencia Regional Faculty (UTN-FRRe). The model proposed in the report of Martínez et al. (2017) is used in this article to follow up the performance analysis with the aim of carrying out predictions with the data obtained during the years 2013 to 2016.

PRELIMINARY CONCEPTS

Data Mining (DM) implies the use of a number of algorithms related to the Artificial Intelligence (AI) field, in an attempt to find regularities in large amounts of information. It mainly aims to search in data, find and identify patterns that allow one to establish reciprocal relations. Once the analysis is carried out, the patterns are used to explain the relations among the data and eventually aid in the decision making process (Han et al., 2012).

Even if the use of these techniques is extensive, the DM process is complex and requires advanced technical and mathematical concepts to be handled. In spite of that, it is worth mentioning that this article focuses on the use of some algorithms that permit to determine students' profiles on grounds of socio-economic and specific examinations data rather than on a detailed description of the process. Therefore, some general DM aspects will be mentioned, being the reader able to look into the theme in works like the ones presented in (Han et al., 2012; Liu, 2011).

The DM process plays an important role in the knowledge discovery in big loads of data. Generally, the stages in DM include (Han et al., 2012): a) initial data cleansing in an attempt to get rid of inconsistencies; b) data integration to unify multiple sources; c) selection of relevant data for the analysis task; d) transformation of data which tend to obtain summaries or added data, the mining process itself that implies applying intelligent methods to obtain data patterns; e) evaluation of patterns obtained to identify patterns of interest; and finally f) presentation of the knowledge obtained by means of visualization techniques and representation of knowledge in a way that the user can understand. Applying the techniques in every step of the process certainly permits the user to finally review and interpret the relations existing in the data set. This analysis is highly useful when making decisions, given that it offers a different perspective about certain situations that usually stay hidden in the data.

Despite the large number of algorithms existing in DM, an efficient implementation requires to analyze what is being searched for and the available data and thereafter from that point create a model. Firstly, the search is based on specific-type patterns and tendencies which are applied to the data set so as to establish some kind of concordance. Supposedly, the acquisition of adequate results is conditioned by the correct choice of algorithms, which is the reason why the decision on what techniques to use turns out to be a crucial task.

Ideally, the algorithms applied are those whose results correlate with the type of data to be treated. In this way, the selection is conditioned by the environment and it is on the expertise of the individual in charge of the process to be able to properly determine what type of algorithms must be used.

The used analysis model implements decision trees and demographic clustering algorithms for the prediction, since besides finding relations among the attributes, groups containing common values are built (Martínez et al., 2017).

Decision trees are supervised learning algorithms (Han et al., 2012; Liu, 2011) which are very popular when finding relations among the values and attributes that describe the data set. They are quite easy to deal with; the objective is to find the nodes which best classify the data. This process is performed recursively and the algorithm adds a node to the tree every time that a significant correlation among the attributes and the classes is found. To establish the convenience to use an attribute as tree node, entropy is used as information gaining measure (Han et al., 2012; Liu, 2011). Throughout the process of tree construction, the total entropy of data set is calculated, and then its entropy respect to each attribute. The node with the largest information gain is chosen in every step of the operation. Information gain is defined as the difference between total entropy and the entropy of the set respect to the attribute.

Being a recursive process, all the possible attribute combinations for each branch are explored, which ends up in a complete tree which, based on the entropy, clearly explains the relations among the attributes.

Although tree representation is useful when analyzing data; the biggest benefit of this algorithm is that it enables to determine classification rules with two associated

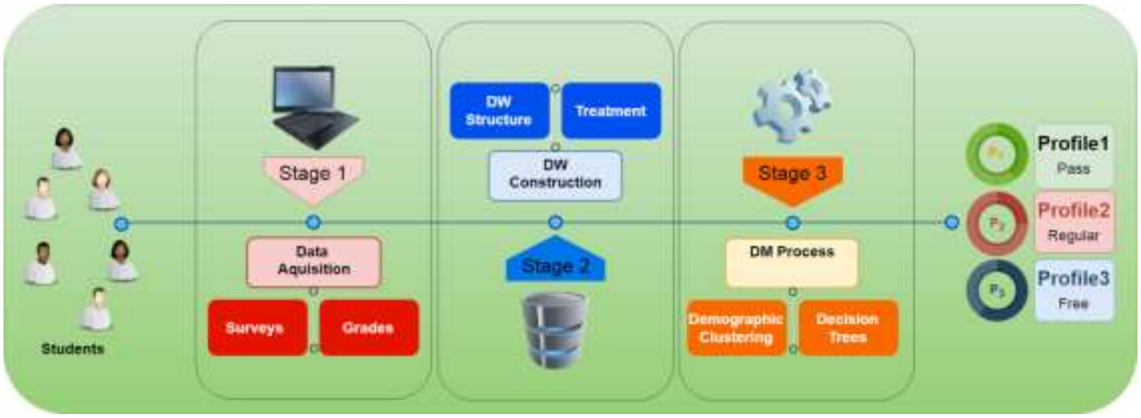

Figure 1: Profile acquisition process.

parameters: support and confidence (Liu, 2011). Support measures the occurrence of the rule in the data set, while confidence measures the reliability that, given certain data, the class predicted is achieved. This kind of algorithm is utterly useful to deal with the problem presented in this article.

On the other hand, the model uses demographic clustering as a grouping technique for data with values of similar attributes. The main idea is the use of distance metrics among the attribute values. Demographic clustering compares each instance to the groups previously established and allots it to that which maximizes similarity valuation. This is an iterative process and performs plenty of data trawling with the aim to minimize assignment error. Cluster quality is measured not only internally (smallest number of reassignments and smallest mean square distance error, etc) (Liu, 2011), but also with intracluster measures (Arbelaitz et al., 2013).

The aforementioned techniques provide the background of the analysis of the proposed model. Their integration is further described.

PROPOSED MODEL

As mentioned earlier, achievement profiles identification plays an essential role both in description as in prediction of special situations of students' performance. In that sense, the proposed model (Martínez et al., 2017) focuses on the use of data analysis techniques in different stages (Figure 1).

The first stage consists of the acquisition of data which will be later considered for establishing achievement profiles. Here, the process aims to collect information about the students related to socio-economic situation and to the grades obtained during the evaluation process. At this point the wide extent of the project can be observed, as it tries to determine which non-academic factors have an influence on students' behaviour, this being the key point in the proposed model.

Socio-economic data are obtained through a survey process which intends to determine certain data that are considered to be relevant to academic achievement. Some aspects to be taken into account among these data could be initially unrelated to achievement, for instance, current accommodation, parents' studies, secondary formation of the student, student's study timing, parents' working situation, student's working situation and even concerns connected to the use of ICTs.

This socio-economic information is collected by means of personal surveys uploaded to digital forms and later stored in a database that, despite having students' identification data are anonymously processed.

Besides, the data acquisition process obtains information related to students' achievement in AED subject during the term. This information consists solely of the grades obtained in each instance of evaluation (examinations and projects assigned for the subject). It is worth mentioning that these grades show achievement in specific situations. Conversely, socio-economic data is depicted daily in general situations.

Data acquisition stage involves both students and professors, being another remarkable aspect, that is, students must provide personal information regarding different social and economic situations and professors must provide the grades obtained along the classes. Consequently, multiplicity of sources of information offers strength to the analysis model.

At the end of term, final situation of the student is added to the database together with their grades. This final situation is divided into three possible values: a) pass, that is, the student has passed the subject with no need to sit for the final examination; b) regular, which means that the student has accomplished the conditions settled during the classes but has to sit for a final examination; and c) free, as for students who have not accomplished minimum requirements for a) or b) and must retake the subject.

The second stage, DW construction, begins when data have been collected (Figure 1). The DW structure proposed by Martínez et al. (2017) is fairly simple as it is made up of

Table 1: Model evolution resorting to students' grades.

Variable	Final situation with 2013-2016 data					
	Without year of admission			With year of admission		
	Without grades	1st Examination	1st and 2nd Examinations	Without grades	1st Examination	1st and 2nd Examinations
Quality						
Model quality	0.696	0.888	0.933	0.596	0.852	0.906
Accuracy for regular student	0.598	0.786	0.930	0.459	0.659	0.930
Classification for regular student	0.608	0.815	0.911	0.459	0.756	0.859
Accuracy for free student	0.956	0.965	0.944	0.972	0.952	0.915
Classification for free student	0.604	0.884	0.954	0.441	0.851	0.922
Accuracy for pass student	0.286	0.821	0.774	0.131	0.905	0.774
Classification for pass student	0.538	0.956	0.941	0.383	0.944	0.935
Confusion matrix						
Correct classifications	79%	90%	92%	75%	87%	91%

only one facts table with several dimensions associated. The facts table contains information inherent to the student and their academic achievement throughout the term. As previously indicated, their final situation is also taken into consideration. The associated dimensions are about aspects to be studied, that is, characteristics to be analyzed (they basically contain descriptive socio-economic information of the student).

In this stage, it is necessary to count on a revision and purification of data, in case of inconsistencies or incompleteness. Basically, these drawbacks come from the proposal of a hardly restricted survey, such that the student can complete the areas in an agile way. Therefore, the data obtained in the previous stage require a purification process consisting of removal or filling null fields and correction of typing mistakes. By this means, consistency and coherence of data are assured.

The last stage (Figure 1) consists of applying DM algorithms to DW data. The demographic clustering algorithms and decision trees earlier described are used to determine achievement profiles. The aim of this stage is to determine which attributes provide a better order to the data set (by using decision trees) and find clusters that explain data characterization according to attribute values (by using demographic clustering). Both techniques allow a dimensional analysis of data by using the final condition of the student (class they belong to) variant as mining parameter.

A general characterization of the three profiles is obtained by the end of the process. This is then used for predicting situations previously unconsidered.

SIMULATIONS AND RESULTS

Test 1: Model validation

The aim of this test is to measure the capability of the model to predict on grounds of a process of cross validation of the data collected during the 2013 to 2016 periods. For that, the model is to be trained with all available data and then predict each one's class as if it were a real classification process. The situation is also analyzed considering and not considering the students admission year and the evolution of accuracy (parameter that measures the probability of correctness of a prediction of the class), the classification (parameter that measures the capability of the model to correctly order records is based on predicted properties), and quality (parameter that measures the general quality of a model, depending on accuracy and classification) resorting to examination grades. Considering all this, obtained results can be observed in Table 1.

It can be observed in Table 1 that model quality increases as students' grades are incorporated, acquiring its maximum value when examination grades are taken into consideration. This behaviour is repeated both for the model in which the students' year of admission is not considered and for the model in which it is. In both cases, quality is more than 0.9, indicating an adequate response of the proposed model. This is a consequence of elevated values for accuracy and classification of each of the categories (Regular, free and pass) that, except for pass accuracy with values greater than 0.9. Figure 2 shows the evolution of model quality for cases in which admission year is considered and cases in which it is not.

The values obtained for the confusion matrix also show that the model performance is adequate when examination grades are incorporated, and this matrix gets to over 90% (Figure 3).

As it can be seen, incorporation of the grades of the first examination highly improves accuracy and classification parameters. However, even though the incorporation of the grades of the second examination improves quality, it does not show big improvement (Table 2).

Table 3 shows the results related to the importance of the fields the model presents. Clearly, the fields that show the grades are highly relevant for the model, both considering

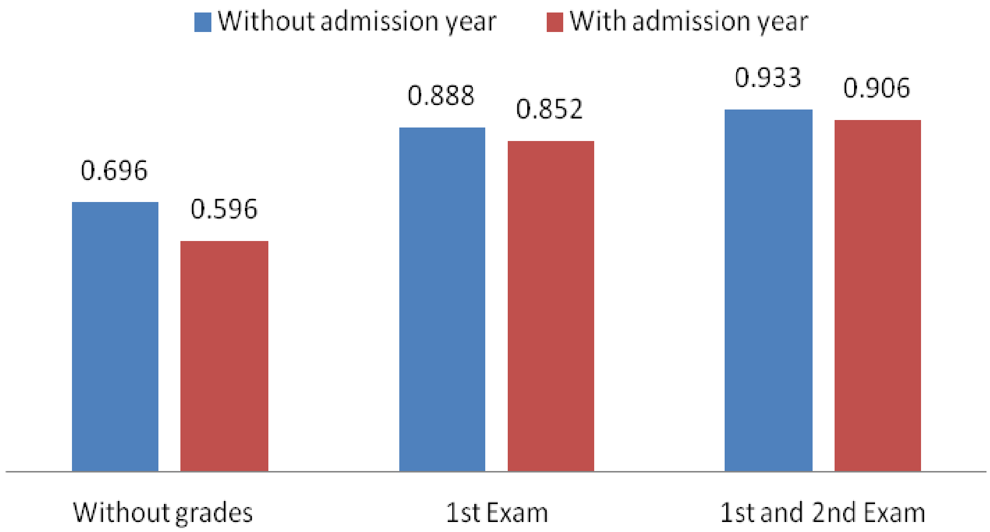

Figure 2: Evolution of model quality considering the incorporation of grades.

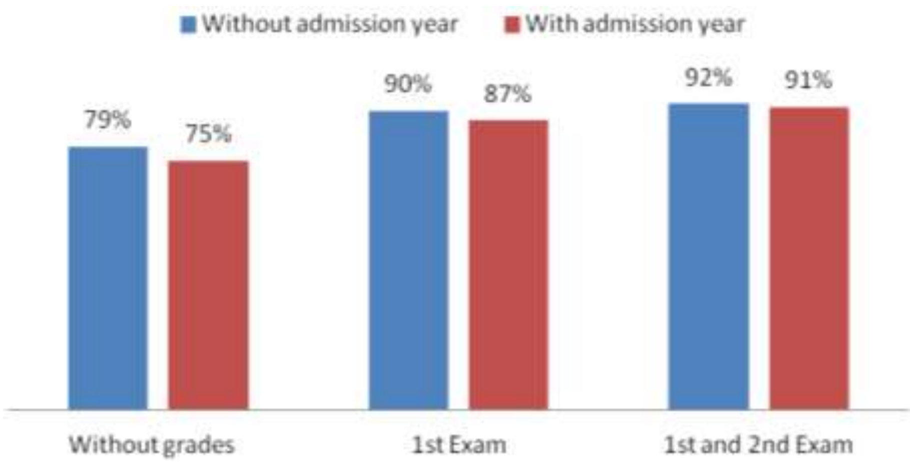

Figure 3: Correct classifications of the model considering the incorporation of the grades.

and not considering the year of admission. In the first case, the grade of the first examination has a 68.49% importance, and as for the second examination, this percentage gets to 77.62%. Nevertheless, when the second examination grade is considered, the importance is drastically reduced, making the field of the second examination grade the most important, with values of 77.46 and 84.69%, respectively.

It is an interesting aspect that in both cases, the most important field not considering the grades is still important in the description, although its importance value decreases substantially.

Test 2: Variation of models considering different years

The aim of this test is to examine the proposed model's behaviour with data from different periods from 2013 to 2015, incorporating data every year. That is, to analyze performance with 2013 data, from the period of 2013 to 2014 data, and with the period 2013, 2014 and 2015 data. Similarly to Test 1, it attempts to measure model quality (by using accuracy and classification parameters).

Table 4 shows that the models in which data from 2013 to 2015 are incorporated behave similarly regarding the incorporation of examination grades of Test 1. Remarkably, from 2013 to 2014 and 2013 to 2015 periods the model quality value decreases substantially if examination grades are not considered. However, they recover adequate quality values by incorporating the first examination grade, reaching quality values greater than 0.9 with the incorporation of the second examination grade.

Figure 4 shows an improvement in the quality of the models for the 2013 to 2014 and 2013 to 2015 periods with the incorporation of the first examination grade observed. By incorporating the second examination grade, quality reaches high quality levels (over 0.900).

Table 2: Quality is improved with the incorporation of grades.

Variable	Final situation with 2013-2016 data			
	Without admission year		With admission year	
	Improves incorporating 1st examination	Improves incorporating 2nd examination	Improves incorporating 1st examination	Improves incorporating 2nd examination
Model quality	0.192	0.045	0.256	0.054

Table 3: Importance of the fields (2013 to 2016 periods).

Without admission year					
Without grades	**Percentage (%)**	**1st Examination**	**Percentage (%)**	**1st and 2nd Examinations**	**Percentage (%)**
Father's occupational category	27.03	First term test grade	68.49	Second term test grade	77.46
Year of Secondary School Graduation	20.53	City of Secondary School	8.79	First Term test Grade	7.12
City of Birth	19.49	Father's Occupational Category	8.07	Year of Secondary School Graduation	5.56
City of Secondary School	12.07	Student's Economic Activity	4.76	Father's Occupational Category	4.38
Mother's Occupational Category	10.09	City of Birth	4.47	City of Secondary School	2.95
Mother's Economic Activity	5.75	Mother's Occupational Category	3.49	Mother's Occupational Category	2.52
Father's Economic Activity	2.72	Father's Economic Activity	1.94		
Father's Study Level	2.32				

With admission year					
Without grades	**Percentage (%)**	**1st Examination**	**Percentage (%)**	**1st and 2nd Examinations**	**Percentage (%)**
City of Secondary School	43.30	First Term test Grade	77.62	Second Term test Grade	84.69
Year of Secondary School Graduation	32.64	City of Secondary School	16.91	First Term test Grade	8.33
Student's Economic Activity	8.58	Student's Economic Activity	3.40	Year of Secondary School Graduation	3.77
Father's Study Level	7.98	Father's Working Hours	2.07	City of Secondary School	3.22
Mother's Economic Activity	3.98				
Importance given to Studying	3.52				

Another remarkable aspect is that, in all cases, with the incorporation of the two grades, confusion matrices with an index of over 90% are obtained, proving the model's strength (Figure 5).

Results related to the importance of the fields present in the model using data from different periods (2013 to 2015), can be observed in Table 5. In the first case (2013 period), year of secondary school graduation is the most important field in the description, even when including examination grades.

For the 2013 to 2014 and 2013 to 2015 periods, the grades field became the most relevant ones for the model, showing similar behaviour to the ones shown in Table 3. In the case of the 2013 to 2014 period, the first examination grade has an importance of 83.88%, and for the 2013 to 2015 period the percentage remains almost the same (83.73%).

When considering the second examination grades, this importance is also reduced considerably like in Table 3, making the field with the second examination grade the most important. Also, in both cases, the most important field not considering the

Table 4: Model variation on grounds of 2013, 2013 to 2014 and 2013 to 2015 data.

Variable	Final situation								
	2013 Data			2013-2014 Data			2013-2015 Data		
	Without grades	1st Examination	1st and 2nd Examinations	Without grades	1st Examination	1st and 2nd Examination	Without Grades	1st Examination	1st and 2nd Examinations
Quality									
Model quality	0.787	0.908	0.904	0.476	0.848	0.904	0.562	0.860	0.928
Accuracy for regular student	0.961	0.843	0.863	0.471	0.826	0.884	0.485	0.819	0.918
Grade for regular student	0.739	0.877	0.857	0.359	0.747	0.864	0.413	0.790	0.899
Accuracy for free student	0.958	0.989	0.958	0.944	0.888	0.944	0.921	0.902	0.943
Grade for free student	0.925	0.984	0.964	0.363	0.869	0.933	0.458	0.893	0.954
Accuracy for pass student	0.208	0.833	0.875	0.000	0.839	0.786	0.095	0.689	0.811
Grade for pass student	0.491	0.873	0.891	0.084	0.921	0.931	0.283	0.911	0.932
Confusion matrix									
Correct classifications	81%	90%	90%	63%	86%	90%	70%	85%	92%

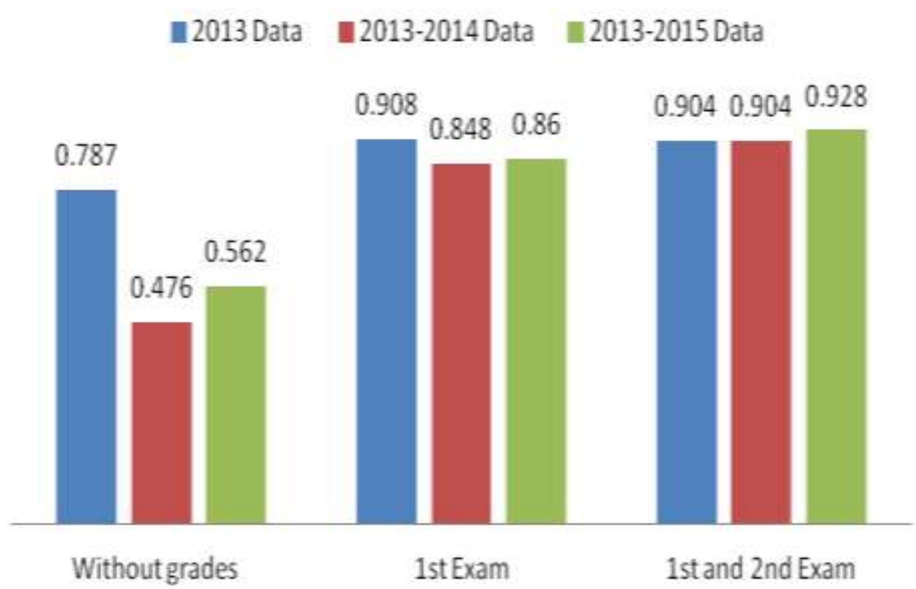

Figure 4: Evolution of model quality using data from different periods.

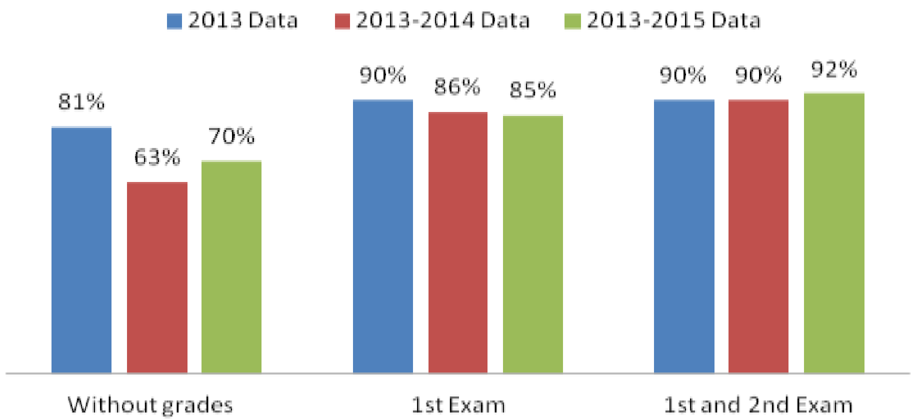

Figure 5: Correct classifications of all the models using data of different periods.

Table 5: Fields importance (variation periods 2013 to 2015 data).

2013 Data					
Without grades	**Percentage (%)**	**1st Examination**	**Percentage (%)**	**1st and 2nd Examinations**	**Percentage (%)**
Year of secondary school graduation	76.86	Year of secondary school graduation	58.07	Year of secondary school graduation	60.40
Father's city of residence	16.91	First term test grade	33.61	Second term test grade	34.63
Student's economic activity	6.23	Student's economic activity	5.14	First term test grade	4.97
		Father's city of residence	3.18		
2013-2014 Data					
Without grades	**Percentage (%)**	**1st Examination**	**Percentage (%)**	**1st and 2nd Examination**	**Percentage (%)**
City of secondary school	61.28	First term test grade	83.88	Second term test grade	82.06
Student's economic activity	38.72	City of secondary school	10.11	First term test grade	14.02
		Student's economic activity	6.01	City of secondary school	3.93
2013-2015 Data					
Without grades	**Percentage (%)**	**1st Examination**	**Percentage (%)**	**1st and 2nd Examination**	**Percentage (%)**
Year of secondary school graduation	61.48	First term test grade	83.73	Second term test grade	78.24
Student's economic activity	13.41	Year of secondary school graduation	4.32	First term test grade	9.69
Mother's study level	10.91	Mother's study level	3.92	Year of secondary school graduation	6.08
Father's economic activity	8.43	Mother's economic activity	3.67	Father's economic activity	4.36
Weekly study hours	5.78	Student's economic activity	2.52	Mother's working hours	1.62
		Mother's working situation	1.84		

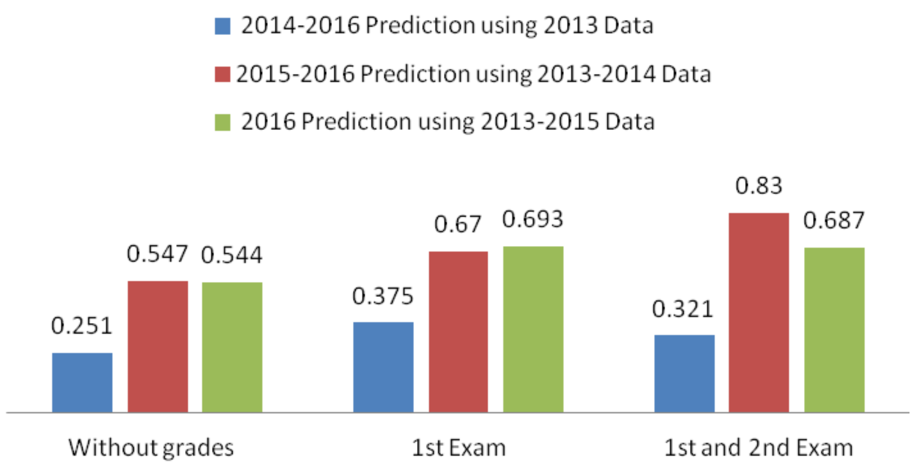

Figure 6: Model quality in prediction of cases not used during training.

grades is still relevant in the description, even though its value substantially decreases.

Test 3: Predictive model based on variation of the 2013 to 2015 periods data

The aim of this test is to use the following configurations in order to measure the quality of prediction of data which have not been used in the training period:

1) Train the model with data from year 2013 and measure the quality of prediction of the data from the period of 2014 to 2016;
2) Train the model with data from the 2013 to 2014 period and measure the quality of the prediction of data from the period 2015 to 2016;
3) Train the model with data from the 2013 to 2015 period and measure the quality of the prediction of data from 2016.

All the aforementioned variants are taken with data not considering grades, but considering the first examination grades and finally, the second examination grades.

Quality values of the model predicting cases in which it has not been trained can be observed in Table 6. An interesting aspect is that the quality is very low if only 2013 data are used for predicting. This occurs due to the fact that the number of data used during training is insufficient to relate students' profiles to characteristic attributes of each class, and it generates very low values of accuracy and classification parameters (some even being negative).

With incorporation of 2014 data, the model substantially improves the prediction quality; reaching a high value when 1st and 2nd examination grades are taken into consideration (0.830). It occurs same with the model trained from 2013 to 2015 data, but in this case, the behavior is repeated in cases where grades are not considered and in which the 1st examination grade is considered. Quality falls to 0.687 with the incorporation of the 2nd examination grade. This behaviour can be clearly observed in Figure 6.

The same behaviour is seen in values obtained for the confusion matrix in which the model trained with data from the periods 2013 to 2014 and 2013 to 2015 is significantly superior to the model trained only with data from 2013. In both cases, by means of the incorporation of the two grades, matrices with indexes of 87% for the 2013 to 2014 period and 83% for the 2013 to 2015 period are obtained (Figure 7).

CONCLUSIONS

This work presented the results of the use of an academic achievement predictive model. It is based on socio-economic and attitude information, as well as, on the results of the first evaluations of students of AED-Algoritmos y Estructuras de Datos (Algorithms and Data Structures) at National Technological University, Resistencia Regional Faculty (UTN-FRRe); it is the continuation of previous work referred to determining characteristic profiles of different levels of students' academic achievement.

The model works on two levels of data, the first one related to socio-economic and attitude information, which is collected through a survey among the students at the beginning of the classes, by means of a computer system developed *ad hoc*. The second level of data corresponds to the grades obtained by the students during evaluations and during classes, and come from the institutional academic system.

On the one hand, the predictive model can be used only with the information that comes from the student surveys performed at the beginning of the classes, that is, with socio-economic and attitude information. This allows one to

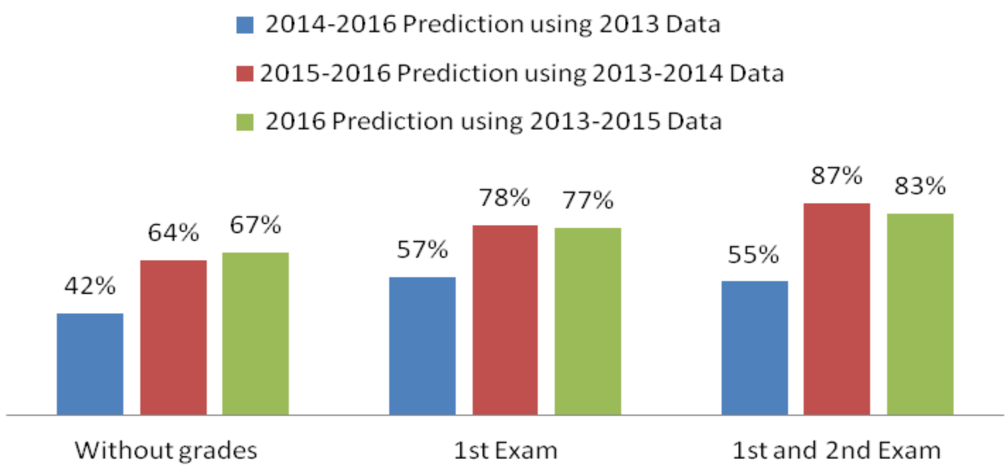

Figure 7: Correct classifications in cases not used during model training.

obtain an early estimate of the academic achievement that students would reach. The results of the model quality and accuracy not considering grades show that the survey is an adequate mechanism for the prediction. In this way, actions specially directed to the group considering academic risk can be planned, and a bigger effort can be made to focus on coaching and special classes, etc, to prevent the group from ending up in academic failure.

On the other hand, the predictive model can also be used by adding academic information, expressed in the grades obtained by the students in their first evaluations, together with the socio-economic and attitude information. With this, the predictive prediction of the model improves and identifies students in academic failure risk along the classes. In this sense, the tests which have been performed show that the quality of obtained results is better, and the accuracy reaches very high levels. Also, by reviewing available studies on predictive models, simulations performed show that the model quality and accuracy increase with the incorporation of bigger quantities of information. Likewise, it is worth mentioning that the model accuracy to predict the free student category is generally superior than that for predicting the regular and pass student categories, which is worthy, given that the free student category is one with the biggest interest, as it implies academic failure.

Finally, it is worth mentioning that even more important than the predictive model itself, is the methodology used for its development, which can be used for generating academic achievement predictive models for other subjects, also taking account of the model being adjustable and improvable by adding more data of the students, allowing the model to be adequate in the long run.

ACKNOWLEDGEMENTS

This work is part of the project "Diseño de un Modelo Predictivo de Rendimiento Académico Mediante la Utilización de Minería de Datos" (Design of an Academic Achievement Predictive Model Through the Use of Data Mining), (Code UTI3808TC) of National Technological University, Resistencia Regional Faculty (UTN-FRRe).

REFERENCES

Arbelaitz O, Gurrutxaga I, Muguerza J, Pérez JM, Perona I (2013). An extensive comparative study of cluster validity indices. Pattern Recognition. 46(1): 243–256. https://doi.org/10.1016/J.PATCOG.2012.07.021

Cerretani PI, Iturrioz EB, Garay PB (2016). Use of information and communications technology, academic performance and psychosocial distress in university students. Computers in Human Behavior, 56: 119–126. https://doi.org/10.1016/J.CHB.2015.11.026

De Moortel K, Crispeels T (2018). International university-university technology transfer: Strategic management framework. Technological Forecasting and Social Change. 135: 145–155. https://doi.org/10.1016/J.TECHFORE.2018.05.002

Di Gresia L (2007). Rendimiento Académico Universitario. Universidad Nacional de la Plata - Argentina.

Dika A, Hamiti M (2011). Challenges of implementing the ethics through the use of information technologies in the university. Procedia - Social and Behavioral Sciences. 15: 1110–1114. https://doi.org/10.1016/j.sbspro.2011.03.247

García M, San Segundo MJ (2001). El rendimiento académico en el primer curso universitario. X Jornadas de Economía de La Educación.

Gökalp M (2010). A study on the effects of information technologies on university students. Procedia - Social and Behavioral Sciences. 9: 501–506. https://doi.org/10.1016/J.SBSPRO.2010.12.187

Hamidi H, Jahanshaheefard M (2018). Essential Factors for the Application of Education Information System Using Mobile Learning: A case study of students of the university of technology. Telematics and Informatics. https://doi.org/10.1016/J.TELE.2018.10.002

Han J, Kamber M, Pei J (2012). Data mining concepts and techniques, third edition. Waltham, Mass. Morgan Kaufmann Publishers. Retrieved from http://www.amazon.de/Data-Mining-Concepts-Techniques-Management/dp/0123814790/ref=tmm_hrd_title_0?ie=UTF8&qid=1366039033&sr=1-1

Jan AU, Contreras V (2011). Technology acceptance model for the use of information technology in universities. Computers in Human Behavior. 27(2): 845–851. https://doi.org/10.1016/J.CHB.2010.11.009

Jiménez MG, Izquierdo JM, Blanco AJ (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. Psicothema 12: 248–252.

Kerimbaeva BT, Berkimbaev KM, Nyshanova ST, Meirbekova GP (2014). The Use of Information Technologies in the Training of Students. Procedia - Social and Behavioral Sciences. 116: 2697–2701. https://doi.org/10.1016/J.SBSPRO.2014.01.638

Liu B (2011). Web Data Mining. Computer Knowledge and Technology (Academic Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-19460-3.

Maletic JI, Collard ML, Marcus A (2002). Source code files as structured documents. In Proceedings 10th International Workshop on Program Comprehension. pp. 289–292. https://doi.org/10.1109/WPC.2002.1021351

Martínez DLLR, Giovannini ME, Molinas MEB, Torre JI, Yaccuzzi N (2017). Academic performance problems: A predictive data mining-based model. Acad. J. Educ. Res. 5(4): 61–75.

Molina JMV, Nicolás MF, et al (2004). Estudio del rendimiento académico universitario basado en curvas ROC. Revista de Investigación Educativa. 22(2): 327–340.

Oloriz M, Lucchini ML, Ferrero E (2007). Relación entre el Rendimiento Académico de los Ingresantes en Carreras de Ingeniería y el Abandono de los Estudios Universitarios.

Spiegel A, Rodríguez G (2016). Students at University have Mobile Technologies. Do they do m-learning? Procedia - Social and Behavioral Sciences. 217: 846–850. https://doi.org/10.1016/J.SBSPRO.2016.02.006

Tinto V (1993). Leaving college: rethinking the causes and cures of student attrition (2nd ed). Chicago ; London : University of Chicago Press.

Cite this article as:

Martinez DLLR, Karanik M, Giovannini ME (2018). Academic achievement profiles: An intelligent predictive model based on data mining. Acad. J. Edu. Res. 6(12): 279-289.

Submit your manuscript at
http://www.academiapublishing.org/journals/ajer