

Predicting Academic Performance of Engineering Students after Approving a Mathematics Leveling Course using Decision Trees

Silvia B. González-Brambila, Lourdes Sánchez-Guerrero, Irma Ardón-Pulido,
Josué Figueroa-González, Beatriz A. González-Beltrán

Universidad Autónoma Metropolitana, Unidad Azcapotzalco, Mexico City, Mexico
{sgb,lsg,ifap,jfgo,bgonzalez}@azc.uam.mx

Abstract. Academic performance of recently accepted students is one of the main issues in Higher Level Institutions since first scholar periods trend to be the most difficult ones for students. Some institutions offer leveling courses to develop students basic knowledge for later courses. However, it is not clear if these help students in more advanced courses. This work presents an analysis, using decision trees, for predicting marks in two mathematics courses based on different criteria of the performance on a previous leveling course. This allows finding the factors that impact in the marks obtained in posterior courses and determining if the leveling one is helping students to improve their academic performance.

Keywords: classification techniques, decision trees, educational data mining, predictive techniques, students performance.

1 Introduction

Data Mining (DM) is defined as the science of analyzing big volumes of data for finding interesting patterns which can lead to knowledge about certain aspects or phenomena [4]. DM uses concepts of machine learning and statistic and considers different kinds of analysis: clustering, associative and predictive (which can be divided into numerical and categorical). Since its appearance, DM has been applied in several areas such as medicine, commerce, finance, business, etc. From year 2000, the concepts and algorithms of DM have been applied in an educative environment, analyzing many aspects related with education, this branch of DM is known as Educational Data Mining (EDM) [8].

One of the most common problems studied in EDM is the academic performance of students, specially using classification or predictive techniques.

The academic performance of students is one of the main issues in Higher Level Institutions (HLI), especially in the recently admitted ones. Transition from High School to College has a great impact in students. Moreover, sometimes the knowledge acquired is not enough to face complex courses and students fail in their first courses. Considering this, HLI implements some leveling courses

whose goal is to give students foundational knowledge for taking more complex subjects.

In the Mexican Universidad Autónoma Metropolitana Azcapotzalco (UAM-A), since year 2008, in all the study plans of the ten engineering programs, a leveling course called Mathematics Workshop (MW) [9] has been implemented whose goal is to level students in basic mathematics concepts so they can have a better performance in posterior and more complex courses. At UAM-A, programs offered at the bachelor level are twelve trimesters (four years of full time studies); then the existence of MW course increases the time of a student for finishing their studies, at least in one trimester.

For approving MW, all admitted students present a diagnosis exam. If their marks are good enough (8 of 10), they do not have to take it and can access to the next mathematics courses Complements of Mathematics (CM) and Introduction to Calculus (IC) which is part of the General Branch Level at UAM-A [10], so all engineering students must take them. If their marks are not good enough, then, students must take MW, expecting that at the moment of approving it, they have the adequate level for taking the next courses.

However, it is not clear whether MW is helping students to have a better performance in CM and ICS. Several students need more than one trimester for approving it, and once they approve it, they do not have a good performance in CM and IC. For this reason, an analysis is necessary which allows us to classify or predict the students' marks in CM and IC, from their performance in MW, and finding the most important factors in students performance in CM and IC.

The structure of the paper is as follows: Section 2, gives a general description of the predictive technique decision trees. Section 3 presents works that have used predictive analysis for studying academic performance. Section 4 contains the steps followed for processing data and obtaining the classification models. Section 5 shows obtained results and models and their analysis. Finally, Section 6 contains conclusions and future works.

2 Predictive Techniques and Decision Trees

2.1 Predictive Techniques

Predictive, also called classification or supervised learning techniques, considers a learning scheme where a set of data is divided in two subgroups; one with a certain classification for teaching or training (called training data) which according to the technique and algorithm used, generates a model that can predict or classify data of the second group, a set of non classified data. Common techniques involved in predicting or classifying data are: decision trees, neural networks and Naïve Bayes classifier. Predictive techniques consider two kind of criteria, that used for classifying (independent variables) and the one to be classified or predicted (decision). Once a model is generated using a predictive technique, its efficiency is tested considering a set of data (test data) which already has a classification. The model must classify this data correctly, assigning

to it the correct value to the decision variable. The percentage of correct classified cases represents the level of efficiency of the model. The accuracy of a classifier model is the probability of correctly predicting the class of an instance, normally is estimated in different ways.

2.2 Decision Trees

Decision trees is one of the most used predictive techniques in DM [12]. As a tree structure is composed by branches and nodes, nodes can be: a root, node or leaf. A leaf represents the assigned classification. Other nodes represent a set of characteristics which lead, through a branch, to other nodes or a leaf. Root is called the best predictor, which is the most important criteria for giving a certain classification. Once the tree is constructed, the predicted or classified value is obtained traveling from the root to a leaf according to the characteristics (independent variables) of the case that is being analyzed.

When a decision tree is generated, the nearer the root is to a node, the more it is relevant for assigning a classification. This means that it has more importance in terms of information quality than other criteria. There are several decision tree algorithms such as ID3 [6], J48 [5] and CART [1].

3 Related Work

Predicting students performance is one of the most studied topics in Educational Data Mining. Several works use prediction techniques for predicting or classifying the performance or marks on exams, courses or scholar periods. In [7], applied predictive techniques over scholar data. They tested at Delhi Technological University data for classifying aspects like: students enrollment preferences, actual demand of certain courses, students that would like to be transferred, satisfaction level and future marks considering several factors. They applied decision trees and artificial neural network techniques. They presented results for enrollment decision and branch predictions showing that C5.0 algorithm of decision trees obtained the best accuracy.

The work of [2] predicted students' performance by using linear regression and matrix factorization approaches. They predicted students' next-term course grades and within-class assessment performance. In particular, they investigated four methods: the course-specific regression (CSpR), the personalized linear multi-regression (PLMR) methods, the standard matrix factorization (MF) and the MF method based on factorization machines (FM) to predict the grade that a student would achieve in a specific course. The results showed that PLMR and MF can predict next-term grades with lower error rates than traditional methods. PLMR were also useful for predicting grades on assessments within a traditional class or online course.

In [3], authors present a predictive analysis for the performance of students from Brazilian public schools. They considered several criteria for two stages, the first before students entered school, and academic criteria was added for

the second stage. Data used consisted in registers of students from year 2015 and 2016. The study determined the causes of student failure for both years. Techniques used involved Gradient Boosting Machine (GBM). Results showed that marks and assistance rates are relevant academic criteria, but social criteria, neighborhood, their previous schools and their age also have an important impact in students performance.

4 Decision Trees Generation

4.1 Obtaining Data

For making the analysis, the CRISP-DM methodology [11] was considered and data was gathered from two sources, the General File of Students (AGA from its acronym in Spanish) which contains information about students, the entrance trimester being used, and the historical record of marks (called kardex) at UAM-A which contains the obtained marks for every student in every course they have taken. In particular, for analyzing how the performance in MW can predict the mark in CM and IC, the marks from students that entered since 2008 and that have already passed MW were considered. Initially, 5,181 students were considered. Marks at UAM-A are assigned with letters: MB (Very Good), B (Good), S (Sufficient) and NA (Not Approved) It is considered that the minimum mark for approving the diagnostic exam is B (Good).

4.2 Generating Sets of Data

After assessing the group of students, their information was processed for obtaining sets of data composed by the following criteria (in parenthesis, the name used for processing it):

- The way a student approved MW, through diagnosis exam or taking it, if MW was taken, the total of chances needed for approving it (WAPR)
- The mark obtained in MW (MMW)
- Time (in quarters) passed after approving MW and taking CM and IC (TAMW)
- The mark obtained the first time CM and IC was taken (approved or not) (MCM and MIC)

Two sets of data were generated, one for the relationship between MW and CM with 5,090 students, and another for MW and IC with 4,228. The difference is because not all students that approved MW have taken CM or IC. Classification variables were MCM and MIC for each set of data respectively. Possible values and their meaning for each variable are presented in Table 1.

Initially, the marks considered at UAM-A were used, but the efficiency of the generated models was very poor (less than 30%). The main problem was that several approving marks (MARK_MB, MARK_B and MARK_S) for CM and IC were classified as not approving marks (MARK_NA). For this reason, the values of the classification variables were changed to the approved mark (MARK_A), that considers all the approved and not approved marks. (MARK_NA).

Table 1. Possible values and meaning for each variable of the data set.

Variable	Possible Values	Description
WAPR	EXAM	Approved through diagnosis exam
	FIRST	Approved through taking the course once
	SECOND	Approved taking the course twice
	MORE_TWO	Approved taking the course more than two times
MMW	MB	Approved with MB
	B	Approved with B
	S	Approved with S
TAMW	NONE	Took CM or IC the same trimester it approved MW
	NEXT	Took CM or IC the next trimester it approved MW
	ONE	Took CM or IC one trimester after approving MW
	TWO	Took CM or IC two trimesters after approving MW
	MORE_TWO	Took CM or IC two or more trimesters after approving MW
MCM or MIC	MARK_MB	Approved with MB
	MARK_B	Approved with B
	MARK_S	Approved with S
	MARK_NA	Not Approved

4.3 Trees Generation

Two decision trees were created, one for predicting the mark obtained in CM and another for the one in IC. For both trees 70% of data for training and 30% for testing their accuracy were used. 10 repetitions were performed considering random sets of data for training and testing. Trees were generated using the algorithms CART, ID3 and J48. Average accuracy and standard deviation for each algorithm is presented in the Section Results and Analysis. Also, trees and rules of best accuracy algorithm are shown.

5 Results and Analysis

5.1 Predicting Performance in “Complements of Mathematics”

After 10 repetitions, the average accuracy using CART algorithm was 61.85% with a standard deviation of 1.03. J48 algorithm had an average accuracy of 61.92% with a standard deviation of 1.077. Finally, CART algorithm produced an average accuracy of 61.38% and a standard deviation of 0.96.

Best accuracy using J48 algorithm was 63.19%. Using ID3 algorithm, 62.9% and using CART algorithm, best accuracy was 61.48%. Confusion matrix with the total of correctly and incorrectly classified of the best accuracy model for each algorithm is shown in Table 2.

Table 2. Amount of correct and incorrect predicted marks in “Complements of Mathematics” per algorithm.

	J48		ID3		CART	
	MARK_A	MARK_NA	MARK_A	MARK_NA	MARK_A	MARK_NA
MARK_A	706	203	723	186	714	185
MARK_NA	359	259	381	237	402	226

As Table 2 shows, all the algorithms have good results classifying approved marks correctly, but they failed in the non approving marks. The most efficient algorithm was J48 whose rules are presented in Algorithm 1 and its tree in Figure 1.

Algorithm 1 Rules of J48 generated tree for predicting performance in “Complements of Mathematics”.

```

if MMW == B then
    MCM ← MARK_A
end if
if MMW == MB then
    MCM ← MARK_A
end if
if MMW == S then
    if TAMW == NEXT then
        MCM ← MARK_NA
    end if
    if TAMW == ONE then
        if WAPR == FIRST then
            MCM ← MARK_A
        end if
        if WAPR == SECOND then
            MCM ← MARK_A
        end if
        if WAPR == MORE.TWO then
            MCM ← MARK_NA
        end if
    end if
    if TAMW == MORE.TWO then
        MCM ← MARK_NA
    end if
    if TAMW == TWO then
        MCM ← MARK_A
    end if
end if

```

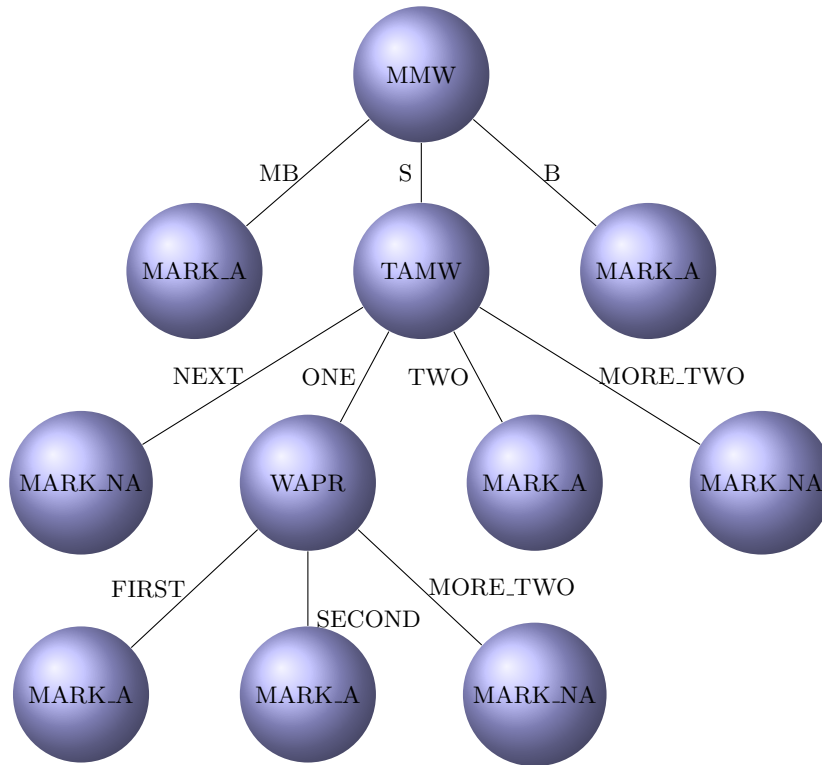


Fig. 1. Decision Tree for predicting the mark in “Complements of Mathematics”.

5.2 Predicting Performance in “Introduction to Calculus”

Average accuracy using CART algorithm was 60.59% with a standard deviation of 1.18. J48 algorithm had an average accuracy of 61.2% with a standard deviation of 0.95. Finally, CART algorithm produced an average accuracy of 60% and a standard deviation of 1.04.

Best accuracy using J48 algorithm was 61%, ID3 60.75% and CART 60.67%. Confusion matrix with the amount of correctly and incorrectly classified instances for the best accuracy of each algorithm is presented in Table 3.

Table 3. Amount of correct and incorrect predicted marks in “Introduction to Calculus” per algorithm.

	J48		ID3		CART	
	MARK_A	MARK_NA	MARK_A	MARK_NA	MARK_A	MARK_NA
MARK_A	268	359	269	358	259	376
MARK_NA	136	506	140	502	123	511

Opposite to CM, the algorithms have a better performance classifying correctly a non approved mark. Similarly with CM, the most efficient algorithm was J48. Its rules for classifying marks in IC are presented in Algorithm 2 and its tree in Figure 2.

Algorithm 2 Rules of J48 generated tree for predicting performance in “Introduction to Calculus”.

```
if MMW == MB then
  MIC ← MARK_A
end if
if MMW == S then
  MIC ← MARK_NA
end if
if MMW == B then
  if WAPR == EXAM then
    MIC ← MARK_A
  end if
  if WAPR == FIRST then
    MIC ← MARK_NA
  end if
  if WAPR == SECOND then
    MIC ← MARK_NA
  end if
  if WAPR == MORE_TWO then
    MIC ← MARK_NA
  end if
end if
```

From the results, it can be seen that the algorithm that obtained the best results was J48 for predicting students performance in both topics. However, the obtained accuracy (63.19% and 61%) is not big enough for determining if MW is really helping or not students in CM and IC. Analyzing the trees, there exists some problems, specially with the CM tree, in particular three branches:

- **S** in **MMW**, **ONE** in **TAMW**, **FIRST** in **WAPR** which leads to an Approved Mark (**MARK_A**)
- **S** in **MMW**, **ONE** in **TAMW**, **SECOND** in **WAPR** which leads to an Approved Mark (**MARK_A**)
- **S** in **MMW**, **TWO** in **TAMW**, which leads to an Approved Mark (**MARK_A**)

It is expected that approving MW with the lowest approving mark (S) and waiting for one or two scholar periods after taking the next courses, will lead to a non approving mark. However, the tree shows the opposite.

Decision trees for predicting performance in IC do not have these kinds of contradictions. Here, approving MW with the lowest mark is related with non approving IC (branch S in MMW). Meanwhile, obtaining the best mark (BM) leads to approve IC (branch BM in MMW).

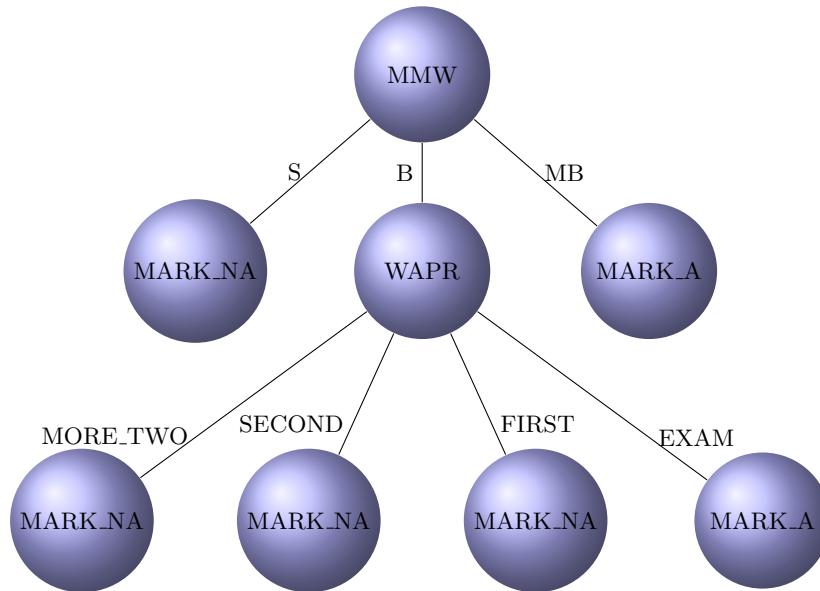


Fig. 2. Decision Tree for predicting the mark in “Introduction to Calculus”.

Besides the confusion matrices, the Sensitivity and Specificity for each algorithm was obtained. Sensitivity and Specificity are two measures commonly used at the moment of predicting and classifying binary cases (with only two possible classified values). Sensitivity or true positive rate measures the proportion of positives cases that are classified as positive. Specificity or true negative rate measures the proportion of negative cases that are classified as negative. Results for each algorithm are presented in Table 4.

Table 4. Sensitivity and Specificity for each algorithm.

	“Complements of Mathematics”		“Introduction to Calculus”	
	Sensitivity	Specificity	Sensitivity	Specificity
J48	0.662	0.560	0.663	0.584
ID3	0.654	0.560	0.657	0.583
CART	0.639	0.549	0.678	0.576

6 Conclusions and Future Work

The aim of this paper was to predict the marks in two mathematics courses according to the performance in a leveling course for determining if it helps students in later courses. However, the obtained models did not have enough

accuracy for being considered reliable. From the analysis, some problems were identified. Having four possible values for assigning in the prediction with decision trees produced a very low accuracy, so, it was necessary to group the approved marks for raising the accuracy. This avoided making a more specific analysis of the obtained marks in “Complements of Mathematics” and “Introduction to Calculus” and was only reduced to determine if the student approved or not. Also, all the algorithms had problems classifying correctly either approved marks (“Introduction to Calculus”) and non approved marks (“Complements of Mathematics”). The Decision tree of “Complements of Mathematics” has some branches which have an unexpected behavior leading to approving marks where it is supposed to expect non approving ones.

Low prediction accuracy could be due to it not being very clear the difference between cases with MARK_A and the ones with MARK_NA, specially in the values of WAPR, MMW and TAMW criteria.

Despite the low accuracy, results from the impact of “Mathematics Workshop” over “Complements of Mathematics” showed that students that took and approved “Mathematics Workshop” with MB or B mark, also approved “Complements of Mathematics” and “Introduction to Calculus”. However, the tree did not give information about the way “Mathematics Workshop” was approved. Students that approved with S and took “Complements of Mathematics” or “Introduction to Calculus” the following period or waited more than two periods, did not approve in their first attempt.

Future approaches for this work include testing other predictive techniques and adding new criteria to the analyzed ones, such as characteristics of the students before entering university. Also, applying some algorithms for leveling the amount of data could improve the accuracy of prediction.

References

1. Breiman, L.: Classification and regression trees. Routledge (2017)
2. Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., Rangwala, H.: Predicting student performance using personalized analytics. *Computer* 49(4), 61–69 (Apr 2016)
3. Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., Van Erven, G.: Educational data mining: Predictive analysis of academic performance of public school students in the capital of brazil. *Journal of Business Research* (2018)
4. Hand, D.J.: Principles of data mining. *Drug safety* 30(7), 621–622 (2007)
5. Patil, T.R., Sherekar, S.: Performance analysis of naive bayes and j48 classification algorithm for data classification. *International Journal of Computer Science and Applications* 6(2), 256–261 (2013)
6. Peng, W., Chen, J., Zhou, H.: An implementation of id3 — decision tree learning algorithm (01 2009)
7. Rajni, J., Malaya, D.B.: Predictive analytics in a higher education context. *IT Professional* 17(4), 24–33 (2015)
8. 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)* 40(6), 601–618 (2010)

9. Universidad Autónoma Metropolitana Azcapotzalco, División de Ciencias Básicas e Ingeniería: http://cbi.azc.uam.mx/es/CBI/Tronco_de_Nivelacion_Academica
10. Universidad Autónoma Metropolitana Azcapotzalco, División de Ciencias Básicas e Ingeniería: http://cbi.azc.uam.mx/es/CBI/Planes_Programa_Estudio.Com
11. Wirth, R., Hipp, J.: Crisp-dm: Towards a standard process model for data mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. pp. 29–39. Citeseer (2000)
12. Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y., et al.: Top 10 algorithms in data mining. Knowledge and information systems 14(1), 1–37 (2008)