



A framework for smart academic guidance using educational data mining

Mohamed Mimis¹ · Mohamed El Hajji^{1,2}  · Youssef Es-saady¹ ·
Abdellah Oued Guejdi¹ · Hassan Douzi¹ · Driss Mammass¹

Received: 21 May 2018 / Accepted: 8 November 2018 / Published online: 26 November 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

The educational recommendation system to provide support for academic guidance and adaptive learning has always been an important issue of research for smart education. A bad guidance can give rise to difficulties in further studies and can be extended to school dropout. This paper explores the potential of Educational Data Mining for academic guidance recommendation by predicting students' performance which involves analyzing data of students' records, socio-economic data and of course the student's motivation. The proposed model was analyzed and tested using student's data collected from the preparatory classes for "Grandes Ecoles" Reda Slaoui (CPGE) - Morocco. More specifically, it proposes the use of three models that were applied on real data: Decision tree, Naive Bayes, and Neural networks. The data include the classes period (2012–2014 and 2013–2015) of 330 students in specialty the grade Mathematical Physics (MP) and Engineering Sciences (MPSI). The performance results indicate that our framework can make more accurate predictions of students' performance.

Keywords Educational data mining · Prediction · Academic guidance · Decision tree · Naive Bayes · Neural networks

1 Introduction

Accurate guidance in education is an essential prerequisite for the success of the academic or professional future of students according to their interests, their academic background, and other criteria related to their personality. Recent studies on this subject

✉ Mohamed El Hajji
m.elhajji@crme fsm.ac.ma

¹ IRF-SIC Laboratory, Ibn Zohr University, Agadir, Morocco

² SADS, Regional Center for Careers Education and Training (CRMEF-SM), Agadir, Morocco

refer to the difficulty to support students in their decision-making (Guichard et al. 2005; Belskaya et al. 2016). Sometimes, they are engaged in courses that do not correspond to their skills, which encourage school dropout. Indeed, students have always been dealing with a serious problem in secondary schools and at universities as well: choosing wisely a specialization that will thereafter define their professional careers. A bad guidance can give rise to difficulties in further study and can be extended to school dropout. Many countries, including Morocco, are offering recourse to advisers to support students in such decision-making situations. However, it is often difficult to give the best choice that matches their profile due to several parameters as well as the grades attained through each course (Castellano et al. 2008). The same issues are emerging also in Morocco 's Classes Préparatoires aux Grandes Ecoles (CPGE) which are considered to be one of the rare educational institutions that prepare Moroccan students to integrate engineering and business schools be it in Morocco or abroad (Belghiti et al. 2017), thus a smart recommendation system to provide support for academic guidance have become a necessity for all professions.

Discovered knowledge can be used to improve student learning and new students' guidance decision-making based on students' profile and predicting the best results. This knowledge would also detect difficulties for students who need personalized interventions in real time. Indeed, a recommendation system of the best choices by predicting their performance with great precision is beneficial because it is possible to identify their academic and vocational orientation at an early stage (Castro et al. 2007). This prediction allows to recognize the real level of the learners, to discover probable risks by intervening to overcome them by offering the best guidance. It allows also to compare between the efforts made and the results achieved and to anticipate the loss or school break. Generally, predicting the future from available observations is an important task, and essential for future decision-making, especially in the field of education, which is difficult and requires many effective techniques to achieve satisfactory results.

New technologies, the Internet, and educational software has created a new context in which a large volume of educational information has still to be fully exploited. Researchers are becoming more interested in data mining techniques in education and the development of the most appropriate algorithms for this kind of data (Castro et al. 2007; Baker and Yacef 2009; Vialardi et al. 2009; Romero and Ventura 2010, 2013). Indeed, a new field called Educational Data Mining (EDM) has emerged as a field of research in recent years. It aims to apply data mining techniques to the educational data (Bakhshinategh et al. 2018), which help to discover many types of knowledge that can help stakeholders to make the best decisions.

This paper presents an effective guidance framework based on educational data mining using predictive models. The objective is to develop guidance prediction models by detecting the most effective factors that work with students' academic performance in the competitions to access the most prestigious institutes and "Grandes Ecoles". The major contributions consist of collecting and annotating dataset which is suitable for smart guidance recommendation, and investigating the suitability of the classifier choice on the accuracy of guidance by analyzing student performance.

The Proposed framework provides a student performance prediction from their database by using educational data mining algorithms. First, the student's data are collected from the preparatory classes for "Grandes Ecoles" Reda Slaoui (CPGE) - Morocco. The data include the classes period (2012–2014 and 2013–2015) of 330

students in specialty the grade Mathematical Physics (MP) and Engineering Sciences (MPSI). This data represents their records, their socio-economic background, and their motivation and inclinations. Then, the predictions are generated by the data mining algorithms: Decision Trees, Neural Networks and Naive Bayes. A system of performance evaluation is adopted to arrive at models that provide the best prediction. Finally, the models obtained are analyzed and interpreted to help counsellors, teachers and students for the best decision-making.

The remainder of the paper is organized as follows: Section 2 presents an overview about educational data mining and different algorithms classifiers used. The architecture framework is presented, the methodology for prediction is specified and CPGE's student profile is defined in section 3. In section 4, we provide the obtained results. Conclusions and highlights on future work are given in section 5 of the paper.

2 Related work

2.1 Educational data mining

As shown in Fig. 1, EDM can be regarded as being the intersection of three main areas: computer science, education, and statistics. This intersection forms also other subareas closely related to EDM, such as computer-based education, data mining (DM), machine learning, and learning analysis (LA).

Each educational data mining technique serves a different purpose depending on the modelling objective. The two most common modelling objectives are: descriptive models and predictive models. In descriptive models, hidden information present in the data is highlighted (e.g., clustering, association rules, sequence discovery, ...) whereas in predictive models, new knowledge is extrapolated from current information (Tan 2006). Various algorithms were used in analyzing students' performance and dropout (Ranjan and Khalil 2008; Tair and El-Halees 2012; Goyal and Vohra 2012; Hung and Zhang 2008). These algorithms are used in deriving a predicted variable (single variable) by extracting models from features.

Ranjan and Khalil (2008) has presented a framework to improve the admission process, delivery of courses in some management schools in order to to improve pedagogy adopted by teachers. The authors concluded that data mining is very helpful for predicting the success of educational programs, and also understanding learning

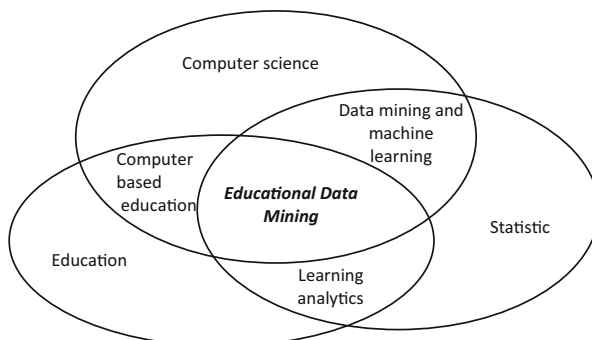


Fig. 1 Educational data mining (Romero and Ventura 2013)

styles in order to promote proactivity in students. Tair and El-Halees (2012) applied EDM to improve the performance of graduate students and to overcome the weaknesses of these students and also tried to extract useful knowledge from graduate student's data collected from the college of Science and Technology – Khanyounis, for fifteen years period [1993–2007]. Goyal and Vohra (2012) examined the effect of the use of EDM models in higher education. The authors applied the EDM techniques to improve the effectiveness of educational institutions. They have shown that the application of these techniques, such as grouping, decision tree and association, at the highest levels of the educational system's processes, can help improve student performance, the life cycle, and the selection of courses. Hung and Zhang (2008) demonstrated how data mining techniques might be utilized to help improve online teaching and learning. Student data was collected from 17,934 server logs of 98 undergraduate students' learning behaviors in an online business course in Taiwan. After a pre-processing of the data, the authors applied techniques of the EDM to discover patterns and knowledge that describe students' behaviour, and to make predictions on learning outcome.

2.2 Algorithms

Decision Trees, Neural Networks and Naive Bayes are the most classifiers commonly used to generate predictive models for student's performance:

2.2.1 Decision trees models

Decision trees models are commonly the most used. It provides a rapid and effective method for classifying and visualizing results. It represents a set of classification rules in a tree form. There are several learning algorithms, for example: ID3 (Iterative Dichotomiser 3), C4.5, CART (Classification and Regression Tree) and CHAID (CHi-squared Automatic Interaction Detector). The basic idea in all algorithms is to partition the attribute space into branches and leaves until the data are classified by satisfying a stopping condition (Brijain et al. 2014). The following table (Table 1) gives a comparative study between these algorithms.

In this study, the C4.5 algorithm has been adopted because the D3, CHAID and Weight-Based algorithms do not support numerical variables.

Table 1 Comparative study between ID3, C4.5 and CART (Brijain et al. 2014)

Algorithm	Data type	Missing Values	Pruning	Connection criterion	Speed
ID3	Categorical	Can't deal with	no	Entropy and information Gain	low
C4.5	Continuous and Categorical	Can't deal with	Pre-pruning	Split info and gain ratio	faster than ID3
CART	Continuous and nominal attributes data	Can deal with	Post-pruning	Gini diversity index	Average

Various study was were proposed to predict and to help identify the rules that can be easily understood by everyone and to develop knowledge. In (Sanjeev and Zytkow 1995), the authors examined the external factors that led to the loyalty of a student at a university by processing the personal database; Student ranking is based on decision tree algorithms. Sacin et al. (2009) proposed an application based on C4.5 algorithm and the deduction of the prediction rules that help the student to make a decision about the choice of courses and the best academic itinerary that allows a great chance to succeed. Decision trees were also used to predict student performance and provide timely recommendations on e-learning systems on the Web (Cakir et al. 2005). To predict dropouts during the first months of university enrollment, Sara et al. (2015) use the CART decision Tree algorithm.

2.2.2 Neural network models

Neural network models (NNM) is a set of interconnected nodes (Hagan et al. 1996). The nodes of a layer are connected to all the nodes of the next level. Figure 2 below shows a sample of a multilayer perceptron architecture as an artificial neural network (ARN) architecture (Yann 1987). This network consists of three layers of neurons: an input layer, an output layer and a hidden layer. The input layer receives an input vector representing the profile to be recognized (corresponding to the attributes), the hidden layer learns to recode the inputs and the output layer represent the classification classes. This layer provides the classification result (Es-Saady 2012).

In the literature, several architectures of neuron networks were also used to predict student's performance. Calvo-Flores et al. (2006) predict student grades from Moodle journals using NNM. The study was conducted on 240 students from the university of Cordoba enrolled in the methodology and programming course. In (Ibrahim and Rusli 2007), artificial neural networks were used to predict the final score of 206 students at a university in Malaysia. A multilayer perceptron topology was also used to predict the likely performance of an intended candidate for admission to a Nigerian university (Oladokun et al. 2008).

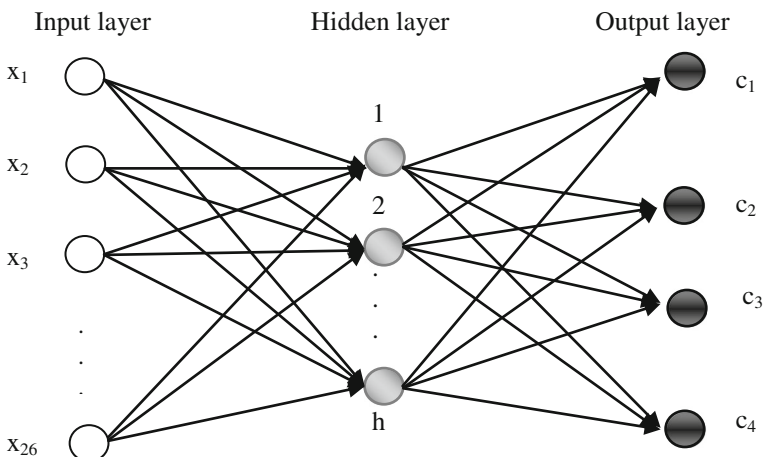


Fig. 2 A multilayer perceptron architecture (three layers: input, hidden and output)

2.2.3 Naive Bayesian model

Bayesian Naive model (BNM) is a classification method based on the Bayes theorem (Wu and Kumar 2009). It is called naive because it simplifies the problem by using two important assumptions: it assumes that the prognostic attributes are conditionally independent, and assumes that there are no hidden attributes that could affect the prediction process. This classifier represents a promising approach to the probabilistic discovery of knowledge and predict students' performance.

The Bayesian classification problem can be formulated using the posterior probabilities: it uses the fact that the value of a predictor X on a given class C is independent of the values of other predictor and assigns to X the class C such that $P(C / X)$ is maximal: $P(C/X) = P(X/C) P(C)/P(X)$.

$P(C / X)$ is the posterior probability of a class C given predictor X .

$P(C)$ is the prior probability of the class C .

$P(X/C)$ is the likelihood of the probability of a predictor X given class C .

$P(X)$ is the prior probability of the predictor X .

Several studies were also proposed to predict student's performance using BNM. Hien and Haddawy (2007) has used Bayesian networks models are to predict student's performance at the Asian Institute of Technology. The Bayesian classifier was used to divide students into groups in order to help teachers to detect students who will need special attention (Bhardwaj and Pal 2012). Oskouei and Askari (2014) studied the students' performance in high school and bachelor degree studies in Iran. They applied various classifiers such as Naive Bayes, C4.5 decision tree, and Neural Networks, and other algorithms, to classify students into 2 classes: Pass, Fail. The results revealed that features such as parent educational level, past examination results and gender impact the prediction.

3 Methodology

The framework is designed for Moroccan preparatory classes for "Grandes Ecoles" (CPGE). Generally, all CPGE institutions provide multidisciplinary programs to develop technical and theoretical skills that allow the student to access the most prestigious "Grandes Ecoles" in Morocco and abroad. It constitutes a post-baccalaureate cycle of two school years. The CPGEs are organized in two main areas: a scientific and technological pole and an economic and commercial pole. The scientific pole is organized into three fields: Mathematics and Physics (MP), Physics and Engineering Sciences (PSI) and Technology and Engineering Sciences (TSI).

After making two or three years in preparatory classes, a student can take the following exam:

- Common National Competition (CNC) for admission to Moroccan engineering high schools ("Grandes Ecoles").
- Competition for access to French high schools ("Grandes Ecoles").

Admission to Moroccan engineering schools is based on the national ranking of students in this competition. Thus, student performance is presented by the rank obtained in the CNC exam (Rang_CNC) that is classified to four classes (M_CNC):

- GA large admitted GA (1st to 300th);
- ADO admitted with oral dispensation (301st to the 1000th);
- MA means allowed (1001st to the 1800th);
- NA not allowed (1801th to the last).

To build a smart recommendation framework that predict CPGE's student performance, and so provide support for academic guidance, we follow the same process of Data Mining; i.e., the same sequence of steps for the extraction of knowledge: data collection, pre-processing, modelling and interpretation. Figure 3 presents the overview of this methodology.

3.1 Data collection and pre-processing

Preparing the variables is often the most important step in predicting problems. Firstly, we created a database containing the information of preparatory classes at the “Grandes Ecoles” of Mohamed Reda Slaoui High School in Agadir, Morocco. For that, we collected the information corresponding to 330 students' grades of the Mathematical

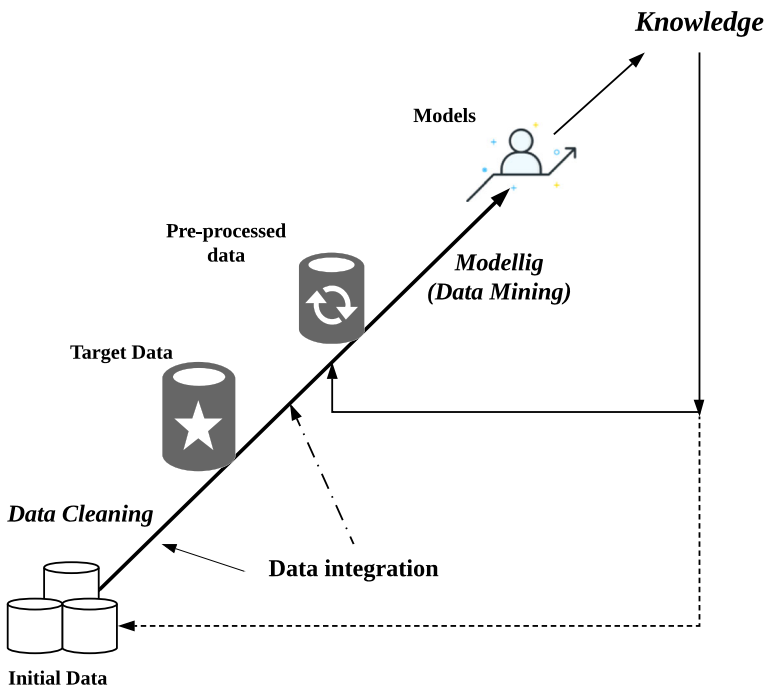


Fig. 3 The methodology used to develop the recommendation framework

Physics and Engineering Sciences (MPSI) courses during the two academic years of 2012–2014 and 2013–2015.

3.1.1 CPGE's student profile and labeling

The only information used to describe the student's profile is a criterion that allows access to the preparatory classes and to support students in the decision-making of their academic career paths. It is represented by the rank of the candidate among all the candidates calculated by the application of the following selection formula (CNIPE - MENFPES 2018):

$$M = N_1 + (N_2 - 10) + \frac{170 \times N_3}{20} + \frac{10 \times N_4}{25}$$

where

- $N_j = 0$ if the student repeats the second year of the baccalaureate, knowing that the Moroccan baccalaureate occurs in 2 years 1st and 2nd year.
- $N_j = 5$ if the student repeats the first year of the baccalaureate,
- $N_1 = 10$ otherwise.
- $N_2 = \frac{M_1 + 2M_2}{3}$, where, M_1 is the general average of the first year of the Baccalaureate, M_2 is the general average of the national baccalaureate.
- N_3 is calculated according to the path chosen as follows:

$$N_3 = \frac{(4M + 3Phy + 0.5LV2) + (1Fr + 0.5Ar)}{9},$$

where M , Phy , $LV2$, Fr and Ar are, respectively, the national examination score of mathematics, physics and English, French and Arabic in baccalaureate.

- N_4 is assigned by the class council of the second year of the baccalaureate (from 0 to 25).

3.1.2 Correlation between the student selection criterion and the final result

As shown in Fig. 4, the correlation between the student selection criterion and the final result at the end of the CPGE curriculum becomes low for the average rankings using the same data of 330 students (110 female and 220 men). It indicates that this selection formula does not represent an effective decision-making tool for guidance and monitoring CPGE students. Indeed, some CPGE students better classified with the selection criterion (R_{sel}) don't maintain the same performance in CNC and reciprocally, some others who are misclassified by the selection formula have good results. Therefore, the current used model must be improved by other characteristics and models that have a better prediction performance.

To further describe the profile of the CPGE student, three types of information were considered. The first concerns the data of selected students before their access to CPGE

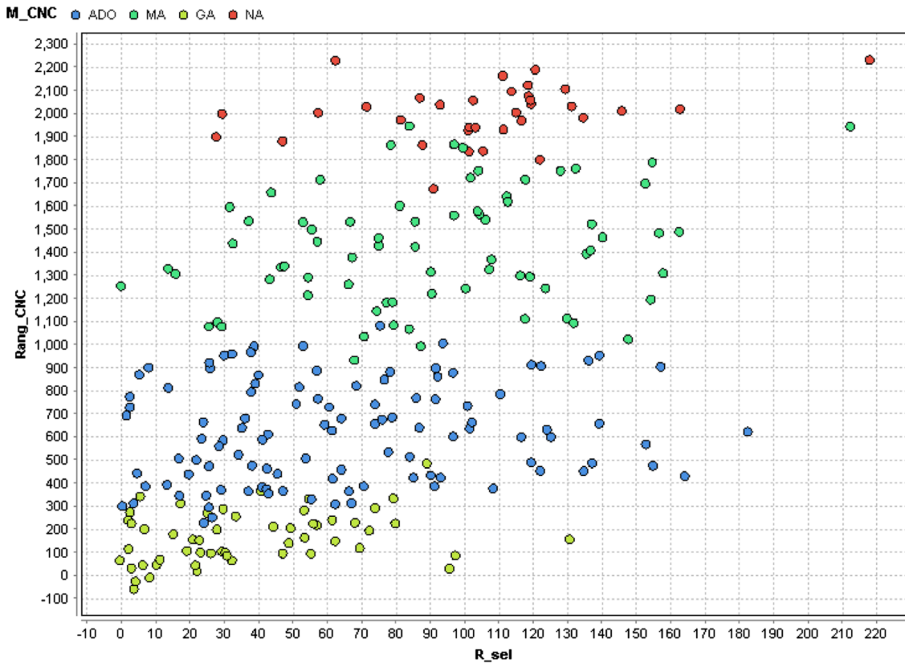


Fig. 4 CPGE students' classification at the end of the school year

(within 2 years of the baccalaureate) presented by the result of the formula of selection. The second type is related to the marks obtained in the first year of preparatory classes. These data concern the general average of student's quarterly marks for each first year for Mathematics, Physics, Engineering Sciences, Computer Science, French, English, Arabic Culture and Translation. Then, the students' scores are transformed in rank; i.e., the student's rank in relation to their classmates.

In order to effectively describe the profile of the student, these data are completed from questionnaires in order to complete the missing data (student's motivations for CPGE Choice, level of parents' studies, number of hours of work outside class) and to improve the prediction models. The introduced variables are described in the following table (Table 2):

3.2 Proposed framework

Figure 5 illustrates the proposed framework; it begins with collection of student's data. Then the models are generated by the data mining algorithms, in particular, the decision tree, Naive Bayes and Neural networks. Finally, a performance evaluation by cross validation is analyzed and interpreted.

This proposed framework can be used as decision-making tools to offer a better guidance and adaptive learning support in order to obtain a higher ranking in the Common National Competition (CNC) based on post-baccalaureate data and student's schooling first year in CPGE.

Table 2 Used variables

Variable	Description
Sexe	Gender (F: female, G: male)
MOT	Motivation of CPGE choice: pa (parents), am (friends) and pe (personal motivation).
MHT	Average weekly hours of work at home (2, 3, 4, 5, 6).
NEP	Father education level: A (none); P (primary); S (secondary); U academic).
R_Sel	Selection rank allowing access to preparatory classes.
RM11, RM12, RM13	Quarterly rankings of the student in Mathematics (first year).
RP11, RP12, RP13	Quarterly rankings of the student in Physics (first year).
RSI11, RSI12, RSI13	Quarterly rankings of the student in Engineering Sciences (first year).
RINF11, RINF12, RINF13	Quarterly rankings of the student in Computer Science (first year).
RT11, RT12, RT13	Quarterly rankings of the student in Translation (Arabic-French) (first year).
RFR11, RFR12, RFR13	Quarterly rankings of the student in French (first year).
RANG11, RANG12, RANG13	Quarterly rankings of the student in English (first year).
M_CNC	Student's score (GA, ADO, MA, NA)

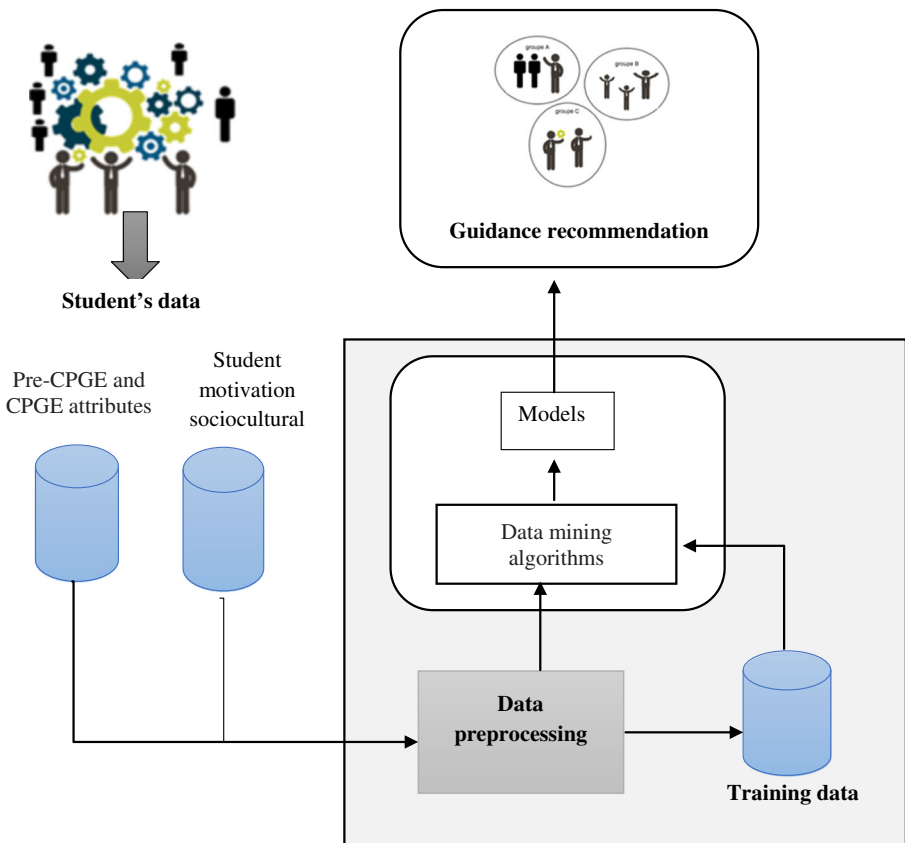


Fig. 5 The proposed framework guidance recommendation

4 Results and discussions

We used three classification algorithms to predict the student's rank in Moroccan common national competition: C4.5 as decision trees algorithm (J48); The neural networks (Multilayer perceptron) and the Naïve bayes classifiers. All those algorithms implementation is available in the Weka collection of machine learning algorithms (Witten et al. 2016).

The validation of the proposed models has been done on samples collected from CPGE database. It corresponds to two groups; namely: Mathematics Physics (MP) and Engineering Sciences (MPSI) during the two academic years of 2012–2014 and 2013–2015. The test data contains a four class (GA, ADO, MA, NA).

The impact of each feature on the score obtained in the common national competition (CNC) is analyzed. For this, we will rank the influence of each one according to a statistical measure. We use the information gain (IG) that determines how well an attribute separates the training data according to the target concept (Baradwaj and Pal 2012). It is based on a measure commonly used in information theory known as entropy.

The entropy is defined as $H(s) = -\sum_{x \in A} p(x) \log_2(p(x))$ where $P(x)$ is the probability of class x when randomly selecting one from the set S .

Information gain $IG(A)$ is the reduction in the entropy that is archived by learning a variable A .

The average information gain and the variance of each feature according to the training data are presented in Fig. 6. The larger value of the information gain is the more influence of the attribute in the induction. The features Sexe and language's attributes don't have a great weight compared to the other variables on the predicted result. It clearly shows that the student's ranking in Mathematics, physics, Engineering Sciences and computer science has a higher rank than social attributes.

The accuracy of three data mining techniques is compared after removing these three variables. 10-fold cross validation scheme is used for result validation (Kohavi 1995). The dataset is divided into 10 subsets, and the holdout method is repeated 10 times. Each time, one of the 10 subsets is used as the test set and the remaining 9 subsets are put together to form a training set. The classification rates of all the 10 trials are averaged to get the prediction result. For the model prediction quality measure, two criteria are adopted: Accuracy and Cohen's Kappa.

In order to get a better insight into the influence of the input variables on the final ranking, we have compared the performance of the model using the pre-CPGE attributes: Sex, NEP, MOT, MHT, R_sel and all attributes. As shown in Table 3, the classification rate increases sharply when students' scores are included in all three quarters of the first year. This shows that the attributes of the student's work during the first year have a significant influence on the prediction of the final result. Indeed, the classification rate for the decision tree algorithm C4.5 is 43.85% when using only pre-CPGE attributes and increases to 54.71% when adding students' scores during the three quarters of the first year.

As shown in Table 3, the accuracy of Naive Bayes and Neural Networks is more important than the interpretation of the model. However, decision trees are easily understandable models. Knowledge models can be directly transformed into a set of rules (if - then) that are one of the most popular forms of knowledge, for their simplicity

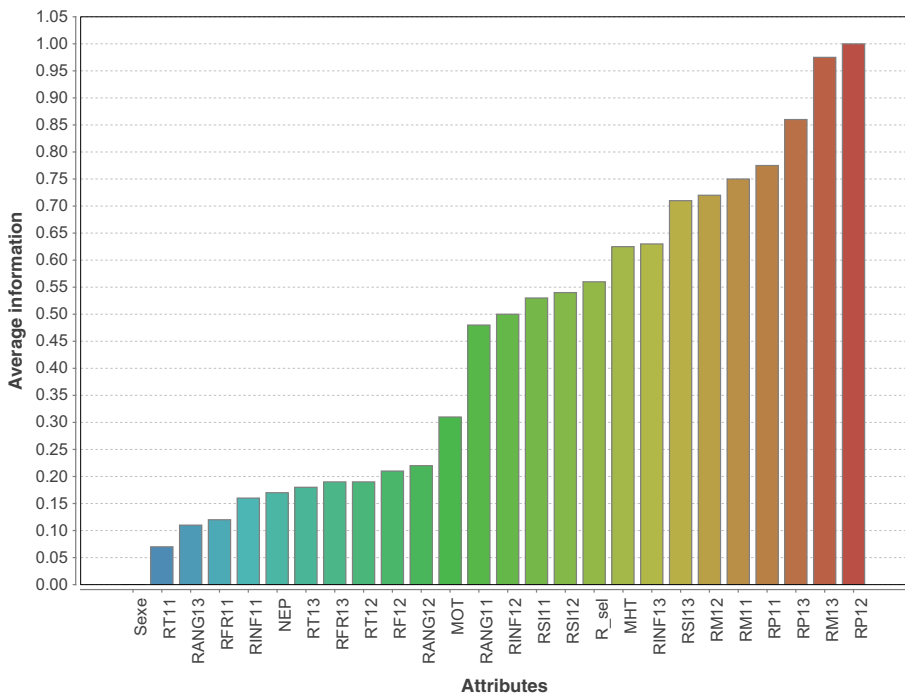


Fig. 6 The average information gain and the variance of each feature according to the training data

and comprehension that teachers or administration advice can easily understand and interpret.

Figure 7 presents a decision tree obtained by the C4.5 algorithm with the parameters: Pruning confidence threshold: $c = 0.25$; Minimum number of instances per sheet: $m = 8.0$. that can be used easily by the guidance council in guidance or the teachers in learning.

Figure 8a presents the confusion matrix for the prediction of the final result using C4.5 algorithm for only the pre-CPGE attributes. From this matrix, we can deduce that the prediction errors are mainly due to the closest classes. The majority errors were made on classes MA, ADO and GA. For example, 39 students from the ADO profile were classified in the profiles corresponding to MA, and conversely, 32 students from the MA profile were classified as ADO that are closer to each other. However, no errors obtained between the two classes NA and GA.

Table 3 Results of the classification for Pre-CPGE attributes and all attributes

Method and algorithm	Pre-CPGE attributes		All attributes	
	Accuracy	Kappa	Accuracy	Kappa
Decision tree (C4.5)	43.85%	0.187	54.71%	0.354
Naive Bayes	49.34%	0.298	58.69%	0.428
Neural Networks (26 neurons for input layer; one hidden layer; output layer of 4 neurons)	52.32%	0.311	56.10%	0.373

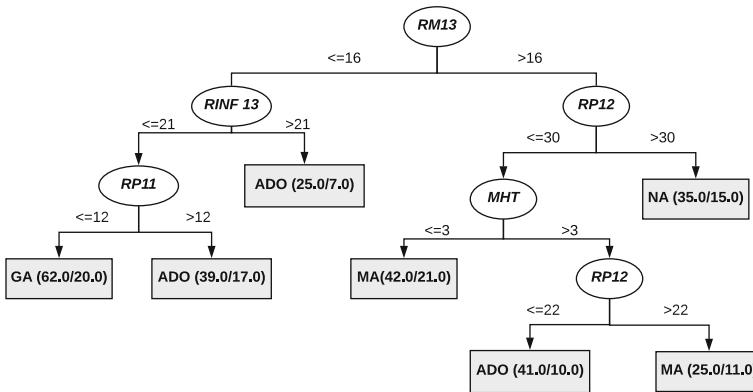


Fig. 7 A decision tree obtained by the C4.5 algorithm: Pruning confidence threshold: $c = 0.25$; Minimum number of instances per sheet: $m = 8.0$

Figure 8b presents the system confusion matrix obtained when all students’ scores of the first year are added. We can see improvements on the errors mentioned above. For example, 22 instead of 39 and 15 instead of 32. This shows that attributes based on the student’s work during the first year have a significant influence on the prediction performance of the final result. For future work, the experiment can be extended with more distinctive attributes to get more accurate results, useful to improve the students’ guidance and learning outcomes. Also, experiments could be done using other data mining algorithms to get a broader approach, and more valuable and accurate outputs.

5 Conclusion

In this paper, we have proposed a framework recommendation to make good decisions for improved learning and guidance. It is based on the prediction of students’ performance using appropriate techniques of the educational. Contrary to what is currently happening in the guidance council or the teachers’ council, which is generally based on assumed main subjects or averages for a fixed period. This framework can make more accurate prediction of students’ performance and guidance. The results show that this performance is dependent on the variables introduced before and during the CPGE training and on the type of the classifier used. The rules generated by the decision tree provide valuable knowledge to guidance counselors, students, teachers in

True:	NA	MA	ADO	GA	True:	NA	MA	ADO	GA
NA:	15	9	2	0	NA:	11	11	5	1
MA:	11	21	32	3	MA:	15	39	15	1
ADO:	8	39	56	25	ADO:	7	22	66	21
GA:	0	4	18	26	GA:	1	1	22	31

Fig. 8 Confusion matrix for the final result using C4.5 algorithm. (a) using only the pre-CPGE attributes, (b) using all attributes

administration in order to make good decisions for the improvement of guidance and learning. A comparative study using different decision-making structures needs to be explored in future work. Also, it consists to generalizing the study to a large dataset in order to train other machine learning algorithms, and to improve the accuracy of prediction of CPGE students' performance.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: a review and future visions. *JEDM| Journal of Educational Data Mining*, 1(1), 3–17.
- Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: a survey of the last 10 years. *Education and Information Technologies*, 23(1), 537–553.
- Baradwaj, B. K. & Pal, S. (2012). Mining educational data to analyze students' performance. arXiv preprint arXiv:1201.3417.
- Belghiti, K., EL Kirat El Allame, Y., & Chana, M. (2017). Critical thinking development: the case of the english course in the CPGE Classes in Meknes, Fes and Kenitra.
- Belskaya, E., Moldovanova, E., Rozhkova, S., Tsvetkova, O., & Chervach, M. (2016). University smart guidance counselling. In *Smart Education and e-Learning 2016* (pp. 39–49): Springer.
- Bhardwaj, B. K. & Pal, S. (2012). Data mining: a prediction for performance improvement using classification. arXiv preprint arXiv:1201.3418.
- Brijain, M., Patel, R., Kushik, M., & Rana, K. (2014). A survey on decision tree algorithm for classification.
- Cakir, M., Xhafa, F., Zhou, N. & Stahl, G. (2005). Thread-based analysis of patterns of collaborative interaction in chat. *AIED*.
- Calvo-Flores, M. D., Galindo, E. G., Jiménez, M. P., & Pineiro, O. P. (2006). Predicting students' marks from Moodle logs using neural network models. *Current Developments in Technology-Assisted Education*, 1(2), 586–590.
- Castellano, E. J., Martínez, L., & Sánchez, P. J. (2008). OriEB, a linguistic CRS for supporting decision making in academic orientation. *World Scientific Proceedings Series on Computer Engineering and Information Science Computational Intelligence in Decision and Control*, Chapter 2. 841–846. https://doi.org/10.1142/9789812799470_0138. Consulted 15/08/2018.
- Castro, F., Vellido, A., Nebot, À., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. *Evolution of Teaching and Learning Paradigms in Intelligent Environment*, 183–221.
- CNIPE -MENFPES (Centre National de l'Innovation Pédagogique et de l'Expérimentation Ministère de l'Education Nationale, de la Formation Professionnelle, de l'Enseignement Supérieur et de la Recherche Scientifique) (2018). Computational simulation of CPGE's selection formula. <http://www.cpege.ac.ma/Cand/Simulation.aspx>. Accessed 23 October 2018.
- Es-Saady, Y. (2012). Contribution au développement d'approches de reconnaissance automatique de caractères imprimés et manuscrits, de textes et de documents Amazighes. Thèse de Doctorat, Université Ibnou Zohr, Faculté des Sciences, Agadir (Maroc). Consulted 15/08/2018.
- Goyal, M., & Vohra, R. (2012). Applications of data mining in higher education. *International Journal of Computer Science*, 9(2), 113.
- Guichard, J., Huteau, M., & Huteau, M. (2005). *L'orientation scolaire et professionnelle*. Paris: Dunod.
- Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural network design* (Vol. 20). Boston: Pws Pub.
- Hien, N. T. N. & Haddawy, P. (2007). A decision support system for evaluating international student applications. *Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual, IEEE*.
- Hung, J.-L., & Zhang, K. (2008). Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching*.
- Ibrahim, Z. & Rusli, D. (2007). Predicting students' academic performance: Comparing artificial neural network, decision tree and linear regression. 21st Annual SAS Malaysia Forum, SAS Kuala Lumpur.

- Kohavi, R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, pp. 1137–1145, Vol. 2): Montreal, Canada.
- Oladokun, V., Adebajo, A., & Charles-Owaba, O. (2008). Predicting students' academic performance using artificial neural network: a case study of an engineering course. *The Pacific Journal of Science and Technology*, 9(1), 72–79.
- Oskouei, R. J., & Askari, M. (2014). Predicting academic performance with applying data mining techniques (generalizing the results of two different case studies). *Computer Engineering and Applications Journal*, 79e88.
- Ranjan, J., & Khalil, S. (2008). Conceptual framework of data mining process in management education in India: an institutional perspective. *Information Technology Journal*, 7(1), 16–23.
- Romero, C., & Ventura, S. (2010). 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.
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27.
- Sacin, C. V., Agapito, J. B., Shafii, L., & Ortigosa (2009). A. Recommendation in higher education using data mining techniques. In *Educational Data Mining 2009*.
- Sanjeev, A. P., & Zytow, J. M. (1995). Discovering enrollment knowledge in university databases. In *Proceeding KDD'95 Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, 246–251.
- Sara, N.-B., Halland, R., Igel, C. & Alstrup, S. (2015). High-school dropout prediction using machine learning: a danish large-scale study. *ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*.
- Tair, M. M. A., & El-Halees, A. M. (2012). Mining educational data to improve students' performance: a case study. *International Journal of Information*, 2(2), 140–146.
- Tan, P.-N. (2006). *Introduction to data mining*. London: Pearson Education India.
- Vialardi, C., Bravo, J., Shafii, L., & Ortigosa, A. (2009) Recommendation in higher education using data mining techniques. In *Proceedings of Second Educational Data Mining Conference*, Córdoba, Spain, pp. 190–199.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann.
- Wu, X., & Kumar, V. (2009). The top ten algorithm in data mining. *International Standard Book*, 13, 978–971.
- Yann, L. (1987). *Modèles connexionnistes de l'apprentissage*. Thèse de Doctorat, Université Pierre et Marie Curie (Paris). <http://www.sudoc.abes.fr/DB=2.1/SRCH?IKT=12&TRM=043586643> Consulted 15/08 /2018.