The 3rd International workshop on Big Data and Business Intelligence (BDBI 2021)
March 23 - 26, 2021, Warsaw, Poland

Artificial Intelligence and Machine Learning to Predict Student Performance during the COVID-19

Ahajjam Tarik[a*], Haidar Aissa[b], Farhaoui Yousef[c]

[a]ahajjam.tarik@gmail.com, L-STI,T-IDMS, University of Moulay Ismail, Faculty of Science and Technics, Errachidia, Morocco
[b]a.haidar@edu.umi.ac.ma, L-STI,T-IDMS, University of Moulay Ismail, Faculty of Science and Technics, Errachidia, Morocco
[c]y.farhaoui@fste.umi.ac.ma, L-STI,T-IDMS, University of Moulay Ismail, Faculty of Science and Technics, Errachidia, Morocco

Abstract

Artificial intelligence is based on algorithms that enable machines to make decisions instead of humans. This technology improves user experiences in a variety of areas. In this paper we discuss an intelligent solution to predict the performance of Moroccan students in the region of Guelmim Oued Noun through a recommendation system using artificial intelligence techniques during the COVID-19.

Keywords: Data Science, artificial intelligent, Machine Learning, prediction, Recommendation, high school, Data Analysis, COVID-19.

1. Introduction

After the Ministry of Education decided to suspend face-to-face studies in Morocco as part of precautionary measures to limit the spread of the Corona virus, the option of distance education was adopted and the rest of the program of Information and counseling and guidance services began to be paid for the benefit of distant students as well. Thus, the various central and regional administrative structures have been involved in this process and the persons in charge of guidance have involved inspectors, advisers, directors of educational establishments and the rest of the interveners to ensure the continuity of the services of distance school, vocational and university guidance in this exceptional circumstance.

* Corresponding author. Tel.: +212 672028620.
 E-mail address: ahajjam.tarik@gmail.com

Before going into this subject in more detail, we would like to recall from the outset that the Ministry of National Education has given, since entering school 2019-2020, the start of the project "Set up an early guidance system and active effective ", and this project is one of the seven projects committed to His Majesty, and it is an ambitious project which aims to improve educational, professional and university guidance, by reviewing its system to monitor and support the learner from the end of primary school in the construction of a personal project, making him An active, pivotal and effective element in the construction of his project and the scrutiny of his academic and professional choices. This project aims to promote the specialized skills that he exercises The educational guidance counselor through group lessons and individual interviews with learners, by setting up the educational support service, and by establishing teaching a space that incubates the learner's personal project, and it also seeks to integrate the guidance component into the school's project as a framework guiding the efforts of all Actors and stakeholders in educational establishments.

Educational guidance is a key to governance and a necessary tool to raise the level of qualification of the human element and to rationalize choices by placing the right person in the right place. It is an ongoing process in the student's school life. And the last months of each school season are marked by an intensive orientation activity which is considered the culmination of the series of awareness campaigns, group meetings and individual interviews that begin at the start of the school season. . Their education or training choices and this is a process that requires learners, families and institutions to participate in a number of processes such as filling out greeting cards, appointment cards for special people and appointment cards for people in vocational training. These processes require studies, arrangements and classification as well as a contract for the councils have pre-screened people and routes with limited access. This period of the academic year is also said to be dynamic at the level of the second year of the baccalaureate, since students, along with the preparation for taking the baccalaureate exams, begin their appointment to enter higher schools and institutes, which they be affiliated or not to universities, with free or limited access, and this appointment requires electronic registration and preparation of the various documents constituting the application files.

The Moroccan education system consists of three levels: primary, secondary and high school; the duration of the latter cycle is three years. It welcomes pupils in the ninth year of basic education who are oriented to continue their studies in a general or technical education section. The age group corresponding to this cycle is between 16 and 18 years of age. This phase leads to the baccalaureate, which opens the way to higher education, or, failing that, access to vocational training.

During the high school level students are led to choose a course of study according to their abilities, preferences and plans for higher education after the baccalaureate, so students are preoccupied with a great intention in their minds because making a wrong choice will influence their personal and professional life.

The role of academic guidance is vital: it is indeed thanks to academic guidance counselors that a student can find his or her career path and future career. Although students are sometimes content to discuss their choices with their families, an outside view can also prove particularly useful!

In the course of their schooling, all children go through phases of orientation that condition their subsequent studies and, to some extent, their working life. A number of reports have just been produced to highlight the main problems involved in providing information and educational and vocational guidance to young people:

- With the multiplicity of streams and branches, the pupil or student no longer knows which stream to choose according to his or her preferences.
- Generally, what is in circulation is that school orientation begins as early as the third year of college, which is wrong: The student must orient himself from primary school.
- It can be said that there is not a good orientation for the student to know his future profession.

Even if there is a good orientation, making the decision and choosing is very difficult, the choice of stream or school will be consistent with the student's preferences and hobbies so as not to fall into other problems such as wasting time.

To this end, a referral system has been implemented that addresses the improvement of the guidance system by predicting student performance in each of the high school streams.

The rest of the article is organized as follows. Section 2 presents some related work; Section 3 presents the data set and proposed methods; and finally, Section 4 provides a conclusion.

In this paper, we use the following notations (see Table 1):

Table 1.

Symbol	Description
IA	Artificial Intelligent
ML	Machine Learning
MSB	Mathematical Science B
MSA	Mathematical Science A
PS	Physical Sciences
LES	Life and Earth Sciences
AMS	Accounting Management Science
AS	Experimental Science
EST	Electrical science and technology
MST	Mechanical science and technology
ES	Experimental Science
MS	Mathematical science
ESS	Economic Science and Management
LSS	Literature and Social Sciences
L	Literature
SS	Social Sciences

2. Related work

Several researches have been carried out in the field of data mining in order to improve the educational system to ensure the good future of the student. Jennifer S Raj [1] presents a comprehensive study of intelligent computation techniques and their applications. Iwin Thanakumar Joseph [2] presents the investigation of data mining algorithms and techniques that could be employed with the intelligent computer system. Fernandes, Maristela, Marcio, Vinicius, Carvalho, and Gustavo [3] They presented a descriptive and preventive statistical study using data mining algorithms to predict the performance of students in the Bronze Capital. Tomasevic, Gvozdenovic, Vranes [4] have done a study to compare machine learning algorithms to predict student performance in exams, i.e. finding students with a "high risk" of dropping out of the course; Uskov , Bakken, Byerly and Shah [5] have proposed a project that evaluates eight ML algorithms to predict students' academic performance in a course; Abu Saa, Al-Emran and Shaalan [6] used the random forest algorithm to predict student academic performance from a new data set from a private university in the United Arab Emirates (UAE); Sekeroglu, Dimililer and Tuncal [7] proposed a system for predicting and classifying student performance using five of the machine learning algorithms respectively; Sossi Safae Alaoui,,Brahim Aksasse and Yousef Farhaoui [8] pollution prediction through IOT and Big data analytics that examines the likelihood to create a fusion between the two new ideas within the context of predicting pollution that happens once harmful substances; Manouselis and Drachsler [9] discuss the importance of technology-enhanced learning; while Thai-Nghe et al. [10] proposed a system that predicted student performance; Thai-Nghe and Drumond [11] proposed a recommendation system that assessed student performance; Romero et al. [12] compared data mining methods to rank students according to their use of moodle; Bekele and Menzel [13] used data mining techniques to predict student performance; Thai-Nghe et al. [14] used data mining techniques to predict students' academic performance; O. Chavarriaga, B. Florian-Gaviria [15] propose a recommendation system that helps students increase their learning skills;AHAJJAM Tarik and FARHAOUI [16] propose a recommendation system for orientation of high school students .

3. Methodology

To address the issue discussed in the introduction, our solution should implement an intelligent system that meets the needs of the students. This will require a system that will predict the student's Baccalaureate average through his

or her core grades using Machine Learning and data mining techniques.

The project therefore consists of the design and implementation of a system for orienting students in the common core towards one of the technical or scientific branches of the first baccalaureate.

The aim of this system is to have a good orientation that will allow the student to get a good grade according to an already existing model that contains the set of students who have already passed their baccalaureate in the region of Guelmim Ouad Noun.

Fig 1 illustrates our approach to predict the pupils' averages, in fact the input data are read from the MYSQL DBMS then transferred into a python script which divides the data into two training and test parts then executes the three algorithms: random forest, decision tree and linear regression and gives at the end the score for each algorithm.
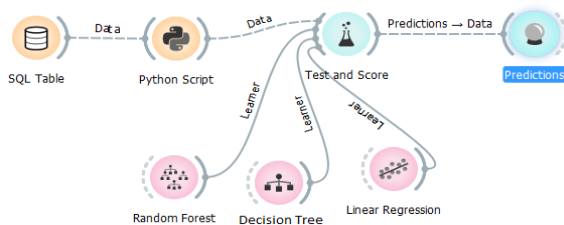

Fig 1. approach of the project

3.1. Data search and pre-processing

For data science, this is the most important step in starting a machine learning project. The data comes from the regional academy of education and training of the region of Guelmim Oued Noun in central Morocco.

Data preprocessing and cleansing are important tasks that must be done before applying machine learning algorithms. Raw data is generally noisy, unreliable, and incomplete. Their use in modeling can lead to misleading results.

Once the data is processed, a series of cleanups will be performed to remove irrelevant items and correct the issue of missing values.

Initially the regional academy of education and training provided us with the grades of 142110 students after data filtering (deletion of missing values and students who did not have their baccalaureate), it remains that the averages of 72010 students enrolled between 2000 and 2015.

The main objective of our research is to have an intelligent system that can predict the average of the baccalaureate from an already existing model.

3.2. Presentation of data

The Gestion_Notes database contains the grades of students in the lycée cycle at three levels: Common Core, first year baccalaureate and second year baccalaureate.

The table below shows the orientation of a student from each branch in the different streams chosen in the first year of his high school career (see Table 2):

Table 2.

	Technical Branches	Scientific branches	branches of Literature
Level I	Technical Core	core science	core Literature
Level II	MST	MS	LSS
	EST	ES	
	ESM		
	MS		
Level III	MST	MSA	L
	EST	MSB	SS
	ES	PS	
	AMS	LES	
	MSB		

3.3. Select model

This step corresponds to the machine learning phase itself of the Data Science project.

It is a matter of choosing the different machine learning models that make it possible to best model the target variable to be explained (business problem).

3.4. Algorithm used

In the realization part we implement 3 algorithms for regression to give a good prediction:
1) Linear regression:

The aim of simple (resp. Multiple) regression is to explain a variable Y using a variable X (resp. Several variables X_1, ..., X_p). The variable Y is called dependent variable, or variable to be explained, and the variables X_j (j = 1, ..., p) are called independent variables, or explanatory variables.

 The theoretical model, formulated in terms of random variables, takes the form of :

$$Y_i = a_0 + a_1X_{i1} + a_2X_{i2} + \ldots + a_PX_{iP} + \varepsilon_i, \qquad i = 1, \ldots, n \qquad (1)$$

2) Decision Tree

A decision tree is a regression and classification algorithm and a supervised decision support tool, structured as its name suggests, like a tree. A decision tree allows a population of individuals to be divided into homogeneous groups according to discriminating attributes based on a fixed and known objective. A decision tree is composed of a root node, a set of internal nodes and a set of leaves. Each internal node of the tree represents an attribute, the external nodes, called leaves, represent the assignment classes, the arcs between the nodes represent the tests on the nodes. The internal nodes of a decision tree have a single parent node, and two or more descendant nodes.

3) Random forest

Decision tree forests or Random Forests are an ensemble learning technique based on decision trees. The random forest model involves the creation of multiple decision trees using split data sets from the original data. And by randomly selecting a subset of variables at each step in the decision tree. The model then selects the mode of all predictions for each decision tree.

The general idea behind the method is the following: instead of trying to obtain an optimized method at once, several predictors are generated before pooling their different predictions.

3.5. Results

At the end of the first year of the lycée (common core), students are led to choose a course of study according to their preferences and aptitudes, also taking into account the studies envisaged after the baccalaureate.

The aim of this system is to have a good orientation which will allow the student to get a good grade according to an already existing model which contains all the students who have already passed their baccalaureate in the region of Guelmim Oued Noun.

After testing the three algorithms we constructed a table that illustrates the score of each algorithm. (see Table 3):

Table 3.

		Data	Algorithm Accuracy %		
			linear regression	Decision Tree	Random Forest
branches of Literature	L	28305	27.81	23.32	54.43
	SS	24290	15.58	12.21	56.87
Science branches	MSA	721	33.62	32.07	61.08
	MSB	241	35.60	23.53	53.66
	PS	8792	42.46	31.21	52.37
	LES	9255	34.70	31.52	70.04
Technical Branches	MST	71	55.45	53.96	54.25
	EST	140	47.80	50.79	54.65
	ES	122	24.16	33.17	60.39
	AMS	52	36 .96	10.69	48.04
	MSB	31	40.16	32.21	46.35

We note that the best model for predicting the baccalaureate average, using a regression algorithm, is the random forest. The average score for this model is higher, because the Random Forest algorithm is composed of several decision trees.

4. Conclusion

In this article, we presented several prediction models. We were able to predict the baccalaureate mean as a function of many explanatory variables (the grades of the core subjects). In order to obtain the best possible results, many models were tested.

Here is the list of the models we have used:

1. Linear regression
2. Regression-type decision tree
3. Regression-type random forest

Finally, we adopted the random forest algorithm which gave us the best predictions than the two other algorithms tested.

References

[1] Raj, Jennifer S. "A COMPREHENSIVE SURVEY ON THE COMPUTATIONAL INTELLIGENCE TECHNIQUES AND ITS APPLICATIONS." Journal of ISMAC 1, no. 03 (2019): 147-159

[2] Joseph, S. Iwin Thanakumar. "SURVEY OF DATA MINING ALGORITHM'S FOR INTELLIGENT COMPUTING SYSTEM." Journal of trends in Computer Science and Smart technology (TCSST) 1, no. 01 (2019): 14-24

[3] Fernandes, Eduardo, Maristela Holanda, Marcio Victorino, Vinicius Borges, Rommel Carvalho, and Gustavo Van Erven. "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil." Journal of Business Research 94 (2019): 335-343

[4] Tomasevic, N.; Gvozdenovic, N.; Vranes, S. An overview and comparison of supervised data mining techniques for student exam performance prediction. Comput. Educ. 2020, 143, 103676.

[5] V. L. Uskov, J. P. Bakken, A. Byerly and A. Shah, "Machine Learning-based Predictive Analytics of Student Academic Performance in STEM Education," 2019 IEEE Global Engineering Education Conference (EDUCON), Dubai, United Arab Emirates, 2019, pp. 1370-1376, doi: 10.1109/EDUCON.2019.8725237.

[6] Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Mining student information system records to predict students' academic performance. In International conference on advanced machine learning technologies and applications (pp. 229–239). Berlin: Springer.

[7] Sekeroglu, B., Dimililer, K., Tuncal, K.: Student performance prediction and classification using machine learning algorithms. In: Accepted to be published in Proceedings of 8th International Conference on Educational and Information Technology, ICEIT 2019 (2019)

[8] Alaoui, S.S.; Aksasse, B.; Farhaoui, Y. Air pollution prediction through internet of things technology and big data analytics. Int. J. Comput. Intell. Stud. 2019, 8, 177.

[9] N. Manouselis, H. Drachsler, R. Vuorikari, H. Hummel, and R. Koper. Recommender Systems in Technology Enhanced Learning. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, Recommender Systems Handbook, pages 387–415. Springer US, 2011.

[10] N. Thai-Nghe, A. Busche, L. Schmidt-Thieme, Improving Academic Performance Prediction by Dealing with Class Imbalance, in: Proceeding of 9th IEEE International Conference on Intelligent Systems Design and Applications (ISDA'09), Pisa, Italy, IEEE Computer Society, 878–883, 2009.

[11] N.Thai-Nghe,L.Drumond,A.Krohn-Grimberghe,L.Schmidt-Thieme,Recommender system for predicting student performance, Procedia Computer Science, 2010,pp 2811-2819

[12] C. Romero, S. Ventura, P.G. Espejo, C. Hervs, Data Mining Algorithms to Classify Students, in: 1st International Conference on Educational Data Mining (EDM'08), Montral, Canada, 8–17, 2008.

[13] R. Bekele, W. Menzel, A Bayesian Approach to Predict Performance of a Student (BAPPS): A Case with Ethiopian Students, in: Artificial Intelligence and Applications, Vienna, Austria, 189–194, 2005.

[14] N. Thai-Nghe, P. Janecek, P. Haddawy, A Comparative Analysis of Techniques for Predicting Academic Performance, in: Proceeding of 37th IEEE Frontiers in Education Conference (FIE'07), Milwaukee, USA, IEEE Xplore, T2G7–T2G12, 2007.

[15] O. Chavarriaga, B. Florian-Gaviria, and O. Solarte, "A recommender system for students based on social knowledge and assessment data of competences," in Open Learning and Teaching in Educational Communities. New York, NY, USA: Springer, 2014.

[16] Tarik, A., & Farhaoui, Y. (2019) Recommender System for Orientation Student BDNT 2019: Big Data and Networks Technologies pp 367-370.