

Predicting the baccalaureate students admission: The influence of teacher and administration

Abdesslam Essayad^{1,*}, My Abdellah Kassimi¹

¹ Laboratory LIMA-UIZ, ENSA Agadir, Morocco

* Corresponding author. Email: hajebmen@gmail.com

Abstract

In recent years, many data has been created about the field of education in Morocco after the introduction of the Education Information System since 2013, which can be used to study the results of the baccalaureate. According to the statistics given by the Moroccan Ministry of Education, the results of the baccalaureate were not more than 60% before 2017, and starting from 2018 the results began to rise due to several changes in the Morocco educational system. However, the number of students who dropped out of the qualifying secondary school was 67,000, with a percentage of 7.4% in 2019-2020. In this paper, the challenge is to create a model using several techniques from machine learning that will enable us to predict the results of the current school year after studying the results of previous years. This will help decision-makers to improve students' performance and explore the important factors that influence the prediction of the school-year end outcomes.

Keywords: Prediction, Linear regression, Machine-learning, Keras.

1. INTRODUCTION

Over the past twenty years, the Moroccan Ministry of National Education has known several reforms: the National Charter for Education and Training from 2000 to 2008, Emergency program from 2009 to 2013 and the framework law 51.17 that codified the strategic vision 2015-2030. The Program for International Student Assessment (PISA 2018) [1] placed Morocco 75th out of 79 countries that participated in an international assessment that evaluates 15-year-old students' abilities in reading, mathematics, and science. The Moroccan students in the 4th year of primary school (CM1) and second year of secondary school ranked among the last five countries in the 2019 edition of the TIMSS ranking, which assesses knowledge in mathematics and science. The evaluation of students is one of the most important axes of previous reforms. We will focus in this research on exploiting the results of the baccalaureate for the last scholar years to predict the results of the current scholar year. We will need to build a better model to predict the results of the baccalaureate students as soon as the results of the first semester of the actual scholar year are issued. The predicted results will be handed over to the concerned departments to find out the students who are in urgent need of school support. In order to enable the administration and teachers to know the students who are the model predicted for their failure, so that the pre-intervention can support them academically, psychologically or socially. The Moroccan baccalaureate includes two levels: the first year so-called regional and baccalaureate second year.

The rest of this paper organized as follows: in the second part, we will address the research reasons. In the third, we will cite the related works. In the fourth, we will refine the dataset and configure it to be usable by the selected algorithms. In the fifth, we will implement the valid models by experimenting with a set of algorithms used in forecasting. We will conclude the paper with notes and suggestions that may be useful in improving the topic search.

2. RESEARCH REASONS

The art of predicting student performance is highly beneficial to everyone especially to the educational administrators and students [2]. Early prediction of the students' academic performance at the end of a training course is one of the challenges in educational centers [3]. By analyzing students' performance, a strategic program can be well planned during their period of studies in an institution [4].

Through the baccalaureate results statistics, it was found that there has been a tangible progress in the results, as it moved from 60% in 2016 to 78% in 2019[5]. This development in the results is due to several reasons, including:

- ✓ Development of an information system «MASSAR» to manage schooling investing results.
- ✓ Management's intention to fight cheating in exams.

- ✓ The framework of the expanded regionalization that has been launched since 2016, giving more powers to the regional academies of education and training to closely monitor the educational issue while encouraging competition between the concerned.
- ✓ The Framework 17.51 adopted nine projects all aimed at improving educational work.
- ✓ The diligence of Moroccan families in educational support for their children
- ✓ Organization of the educational system, where tests and exams are fixed by calendar throughout the school year.
- ✓ Adopting the success rate instead of the transition threshold

The intervention of officials in the academic field in order to raise the level of scholar achievement of students will inevitably lead to an increase in the school results. There are difference in the results between institutions, directorates and academies. For example, the Boulemane directorate of Fez Meknes academy occupied the first nationally over the past three years by more than 90%, despite the fact that the region is rural and vast and is located in the Atlas Mountains, and it lives a period of severe cold. The secret of this, according to official sources, is that there are civil society associations that provide school support to baccalaureate candidates, especially when the exams are approach.

The aim of the research is to help decision makers and enable them to intervene early to support students expected to fail. Teachers can also focus their efforts and identify the students who have material weaknesses in order to address this weakness with support within or outside the classroom. The administration can tell parents that their mission is to help and provide material and psychological support to their children. Likewise, if students know that they are not expected to succeed, they may increase their efforts and perseverance for success. Also for Directories of national education institutions and educators to determine their needs of school support, according to the materials that its points pose a risk to the results in a timely manner to achieve success at the end of the school year.

The ability to predict the educational results of students is very important in the educational system. Recently, machine learning techniques have been applied to analyze educational data focused on retention and graduation rates to develop a more efficient and responsive system to support students.

The national program for assessing learning PNEA2016 [6], focused on the quality of teaching at the national level through targeted questions that concern the first year of the baccalaureate. The study focused on the opinion of the students and their satisfaction with the level

of teaching and supervising, and concluded the ratios of the effect on the learning achievement at the level of administration and the student. Based on the foregoing, the administration's cooperation with teachers and other factors may contribute to raising the percentage of success at the end of the year is about 20%, and the 80% is for the student's efforts [7]. The administration will be able to identify the students whom the model predicted their failure admission, also to interfere in their academic and psychological support, whereas the social and economic support remains the prerogatives of other parties. This approach should enable the secondary institution management to develop an easier educational policy that can support students at risk or increase motivation for students who will get high rates.

Predicting successful and unsuccessful students at an early stage of the degree program help academia not only to concentrate more on the bright students but also to apply more efforts in developing programs for the weaker ones in order to improve their progress while attempting to avoid student dropouts [8].

3. RELATED WORKS

Predicting student performance is an important application of educational data mining, As stated in the paper entitled: "Predicting Critical Courses Affecting Students Performance: A Case Study" by the researchers' Yasmeen Altujjar, Wejdan Altamimi, Isra Al-Turaiki, Muna Al-Razgan [9]. In the study "Review on Predicting Students Graduation Time Using Machine Learning Algorithms" [10] by the researcher Nurafifah Mohammad Suhaimi from Malaysia, where she compared interest's student's performance in terms of gender; she found that the method of Male and female studies are different. In addition, some researchers cared about the age factor, they concluded that the older the age, some physical changes occur in the brain which lead to more difficulties in remembering or learning efficiently. K. Juliani; A. Mewati [11] used the Apriori algorithm to compare students' performances in common courses at the undergraduate and post-graduate levels, they discovered associations, and then identified factors that determined students' chances of success or failure like syllabus plan, student's interest, teaching and evaluation techniques. B. Baradwaj and S. Pa (2011) [12] performed a study using Bayesian classification on bachelor of computer application students from Awadh University in India, the authors found that the student's performance is strongly correlated with other elements more than students' effort such as family income, students' routine and other different' factors as mentioned. Strecht, Cruz, Soares, Merdes-Moreria, and Abren(2015) [13] predict students' success/failure and grade in a course by using social variables like age, sex, marital status, nationality, displaced (whether the student lived outside the district),

scholarship, special needs, type of admission, type of student (regular, mobility, extraordinary), status of student (ordinary, employed, athlete, etc.). ElGamal (2013) [14] predicts students' grades in a programming course by considering different factors like the student's mathematical background. Huang and Fang (2013) [15] predict course performance based on students' performance in prerequisite courses and midterm examinations. Romero, Lopez, Luna, and Ventura (2013) [16] investigated the appropriateness of quantitative, qualitative and social network information about forum usage as well as the appropriateness of classical classification algorithms and clustering algorithms to predict students' success or failure in a course. Arnold and Pistilli(2012) [17] provide an early intervention solution for difficult courses based on students' activity in a Learning Management System. Applications of data mining techniques in educational environments are the most common research areas, and there is an increasing number of researches published in recent years (Minaei-Bidgoli & Punch, 2003; Schumacher et al., 2010; Etchells et al.; 2006; Aisha et al., 2010; Wok et al., 2009) [18]. Research in educational data mining focuses on various aspects of the educational process: students, teachers, teaching aids, organizing teaching, etc. (Falakmasir & Habibi, 2010; Kumar & Chaddha, 2011; Romero and Ventura, 2007; Guru & others, 2010) [19]. Gonzalez and DesJardins (2002) [20] attempted to predict student success, using linear regression models and a neural network; the results of their research indicate that logistic regression models do not predict as much student behavior as models of artificial neural networks. Thomas and Hass (2001) [21] compared the performance of the three different data mining techniques to predict student's behavior: neural networks, mass algorithms and decision trees, with the neural network model achieving the best results. Delgado (2006) [22] used neural networks to predict student success in exams, identified in binary classes (success or failure).

4. DATA PREPARATION AND PROCESSING

Big data processing is a crucial task for many researchers, administrators, organizations and companies to collect the data and analyze the huge amount of specific data or information. Data preparation is highly recommended for many reasons such as dataset quality, process of data analysis, possibility of related algorithms to apply for removing noisy and missing data and increase the data reliability that is high-quality data models require high quality of data [23]. Data processing is transforming the data into a basic form that make it easy to work.

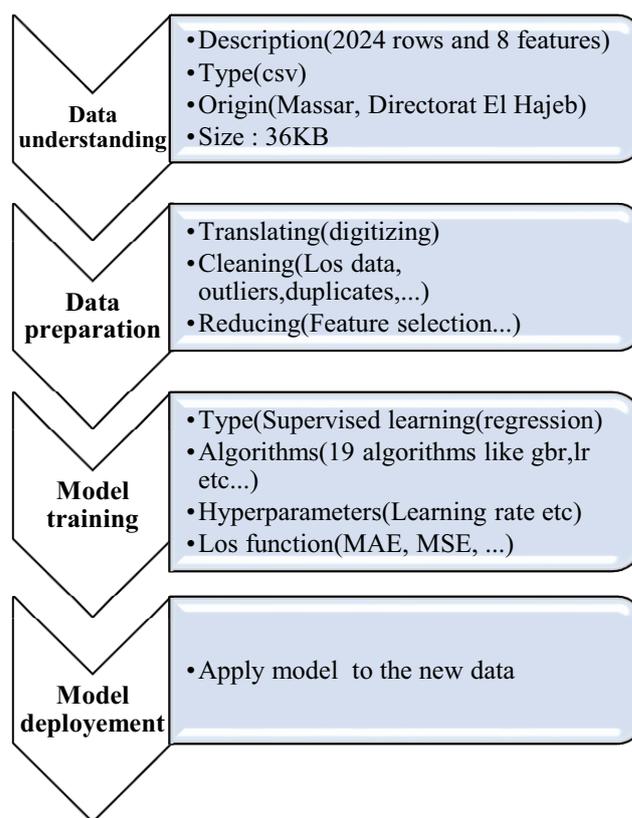


Figure 1 Framework for Data Preparation Techniques

4.1. Data Understanding

For a specific data analysis project, the first step is to understand the data requirements of the project in conjunction with the problem definitions and expected objectives [24].

The main objective is to conduct some experiments on the results of the baccalaureate level using the techniques of data extraction and derive the knowledge that will prove useful to the Directorate of the Ministry of National Education in El Hajeb, which is our field of research in order to make the right decisions necessary to increase the success rate.

The dataset used is the results of the baccalaureate for the previous scholar year; we will use it on a predictive model for the academic results for the first semester of the current scholar year. Two aspects of student performance have been relied upon, namely the regional average for the baccalaureate first year(regional) and the average of continuous monitoring for the first semester of the current scholar baccalaureate second year. We will use the two years' exam dataset, which is collected from the MASSAR system. This dataset has 8 features, 1 target, and 2024 records described below:

Table 1. Dataset description

Variable	Type	Definition
id_student	Integer	
id_school	Integer	
id_milieu	Integer	rural or urban center
id_level	Integer	school level
Sex	Integer	Gender(Male: 1, Female: 2)
nb_day_born	Integer	
reg_average	Number	the regional average(0,20)
cc_average	Number	the average of the continuous control(0,20)
ga_average	Number	General average(0,20)

4.2. Data Preparation

This step is very important for any data initialization process, as it specifies the data to be extracted, including collecting data from the source, preparing it and pre-processing it, so that it is consistent to facilitate the data analysis process. This Step should be taken to carry out before applying analysis and modeling processes. In the case of most educational data, pre-treatment should include the following steps:

Data translation: digitizing records by changing names to numbers, this step is important to understand the data and simplifies the process of using and merging it. We will convert the binary columns into numerical columns (One Hot Encoding).

Data cleaning: standardizing the use of values indicating one thing, for example some institutions store estimates as a letter while others store them as numeric values, also replacing lost data and removing unreasonable, repetitive or contradictory values, as well as filling missing data with reasonable values. Standing in some unusual situations, such as the student getting rates in the national exam, far exceeding the rates of continuous monitoring (outliers' points). Outliers points are the rare values in categorical variables tend to cause over-fitting for the model.

Data Reduction: Retain data that provides knowledge that can be used in the analysis. The explosive growth in the electronic data of qualifying secondary education institutions creates the need to obtain some useful information from this large amount of data. "The process used for transforming raw data compiled by education systems into useful information that could be used by lecturers to take corrective actions and answer research questions" [25]. It is probably better for researchers to

emphasize the specific ways in which variable importance is operationalized [26]. Checking which columns are relevant and dropping irrelevant columns (Feature selection). By performing further analysis, we can conclude that grades in the exam in the first semester of studies, as well as regional average from the Baccalaureate first year. On the other hand, the gender of students and nb_day_born are not of great importance for predicting student success. After that, we will only maintain the regional rate (reg_average) for the first year in baccalaureate and the continuous monitoring (cc_average) for the second year, and the result of the end year (ga_average). We note a correlation between the continuous monitoring rate and the general rate. All students who score in continuous observation have a high pass rate.

Table 2. Correlation between features and target

	cc_average	reg_average	ga_average
cc_average	100%	54%	83%
reg_average	54%	100%	74%
ga_average	83%	74%	100%

Understanding the relationship between variables is useful, we can use the value of one variable to predict the value of the other variable, and the correlation between variables indicates that as one variable changes in value, the other variable tends to change in the same direction. In our case, cc_average had an 83% effect to predict ga_average.

To prepare our dataset to be used very well, there are several techniques for doing that, including the Python Library like Panda, NumPy, and Sklearn, etc.

4.3. Model training and Evaluation

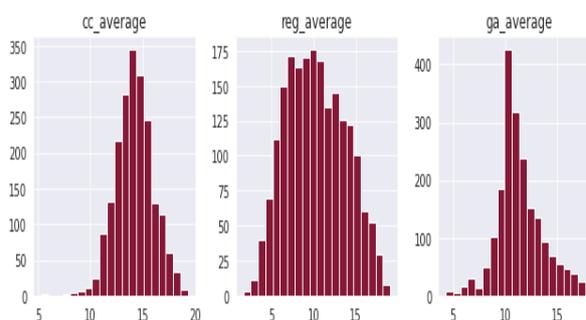


Figure 2. Distribution of features and target values.

In the figure2 all the features have, a normal distribution not skewed distributions; however, we have no problem to use regression model with machine learning algorithms.

There are several factors to consider, when choosing between algorithms, most of them works well with small data sets; deep neural network recommended when we have much data and require time and material to train. The objective is to find the students' results of the school year. Therefore, the problem enters supervised learning and represents a linear regression model. Here are some important considerations while choosing an algorithm like the problem statement and the kind of output we want, the type and size of the data, the available computational time, the number of features, and others observations in the data. We must choose the algorithms and the hyper-parameters, which allow developing the model to make good predictions on new data. Machine learning is a set of methods that make it possible to establish, from data, regressor models for prediction and decision-making. We will work with the various machine-learning algorithms and artificial neural network with the aim of finding the best predictable models for the end of the school year results; that can be identified factors wish have had a decisive impact on students' overall success.

4.4. Deployment

Our main aim is to make a Linear Regression model with two variables, which can give us a good prediction on the final exam. The variables we are basing our predictions on is called regional average and monitoring average variables and is referred to as $X = (x_1, x_2)$. The variable we are predicting is called the general average variable and is referred to as \hat{Y} .

The formula of our linear regression model is : $\hat{Y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \epsilon$

\hat{Y} : predicted output, β : coefficients, x : input, ϵ : error

4.4.1. Deployment with many algorithms from the Scikit learn library

Many machine-learning algorithms that we applied to the dataset using the Pycaret [27] library, show prediction and error rate. We use the 80% for the data to train model for learning and 20% for the data to test model to see if it has learnt the data well or not.

A linear regression model try to find linear relationship between the inputs and outputs. To find this relationship, there exists a method to calculate the optimal coefficients in the formula above, to use with inputs to predict the output. Unless our data is a perfectly straight line, our model will not precisely hit all of our data points. The error metrics will be able to judge the differences between prediction and actual values called by residual; however, we cannot know how much the error has contributed to the discrepancy. In our case, the average of residuals is small, it implies that the model that has a good job at predicting. For building a prediction model, many

experts use gradient boosting regression, it is a machine learning technique for regression and classification problems.

Table 3. The results and performance of the algorithms used by Pycaret.

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT(Sec)
gbr	Gradient Boosting Regressor	0.7297	0.8981	0.9469	0.8189	0.0844	0.0688	0.0390
lr	Linear Regression	0.7597	0.9303	0.9636	0.8125	0.0850	0.0715	0.0210
ridge	Ridge Regression	0.7597	0.9303	0.9636	0.8125	0.0850	0.0715	0.0160
lar	Least Angle Regression	0.7597	0.9303	0.9636	0.8125	0.0850	0.0715	0.0140
br	Bayesian Ridge	0.7597	0.9303	0.9636	0.8125	0.0850	0.0715	0.0130
huber	Huber Regressor	0.7596	0.9313	0.9641	0.8122	0.0851	0.0716	0.0240
catboost	CatBoost Regressor	0.7442	0.9416	0.9694	0.8102	0.0870	0.0705	1.3150
lightgbm	Light Gradient Boosting Machine	0.7735	1.0009	0.9996	0.7982	0.0896	0.0735	0.0270
rf	Random Forest Regressor	0.7853	1.0298	1.0140	0.7926	0.0903	0.0742	0.2220
knn	K Neighbors Regressor	0.7887	1.0279	1.0132	0.7921	0.0900	0.0748	0.0280
ada	AdaBoost Regressor	0.7974	1.0346	1.0157	0.7915	0.0903	0.0753	0.0280
en	Elastic Net	0.8313	1.1214	1.0574	0.7756	0.0928	0.0783	0.0160
et	Extra Trees Regressor	0.8246	1.1153	1.0555	0.7750	0.0940	0.0780	0.1900
xgboost	Extreme Gradient Boosting	0.8150	1.1310	1.0623	0.7724	0.0956	0.0771	0.1570
lasso	Lasso Regression	0.8713	1.2570	1.1191	0.7490	0.0977	0.0820	0.0130
omp	Orthogonal Matching Pursuit	0.9873	1.5128	1.2284	0.6953	0.1074	0.0915	0.0130
dt	Decision Tree Regressor	1.0400	1.7642	1.3255	0.6397	0.1177	0.0973	0.0160
par	Passive Aggressive Regressor	1.0749	1.7979	1.3127	0.6323	0.1088	0.0979	0.0150
llar	Lasso Least Angle Regression	1.7182	5.0506	2.2434	-0.0077	0.1844	0.1594	0.0140

The comparison of each model present in the Table3 depend on the problem statement. Every model has trained using the default hyperparameters and the performance metrics evaluated using the cross-validation.

4.4.1.1. Evaluation Metrics

PyCaret makes it possible to compare models easily. We compare the models by the evaluation metrics (MAE, MSE, RMSE...) using the test data and choice the best predictive result.

Table 4. Evaluation Metrics Description

Metrics name	Usage	Formula
MSE (mean squared error)	Used where we need to highlight large errors. Is most apparent with the presence of outliers in our data.	$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$
RMSE (root mean square error)	Prediction highly affected by outliers, and penalize large errors	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2}$

MAE (Mean Absolute Error)	We use it when we have outliers or we are not too worried about.	$MAE = \frac{1}{N} \sum_{i=1}^N y_i - \hat{y}_i $
R² (R-square)	R ² will always be between -∞ and 1. Negative, a good model have a low R ²	$R^2 = 1 - \frac{MSE(model)}{MSE(baseline)}$
RMSLE (Root Mean Squared Log Error)	The outliers are scaled down, and more penalty where we have prediction less than the actual.	$\sqrt{\frac{1}{n} \sum_{i=1}^n (\log(x_i+1) - \log(y_i+1))^2}$
MAPE (mean absolute percentage error)	We have a large penalty when the residual is negative	$MAPE = \frac{1}{N} \sum_{i=1}^N \left \frac{Z(t) - \hat{Z}(t)}{Z(t)} \right \cdot 100\%$

The output prints a scoring grid with MAE, MSE RMSE, R2, RMSLE, and MAPE, returns the best model based on the metric defined in the sort parameter. In our loss function, MAE is better to use if we need to ignore the outlier's points. However, if we want to account for them, we use MSE/RMSE. Model evaluation mostly based on MSE and RMSE and useful when the outliers are messing with predictions.

Using XGBoost with default hyperparameters give us, 0.81 by MAE loss function as we can see in the table3. However, that gives 0.80 by R2 when used to make predictions on new data as shown in the figure3 below.

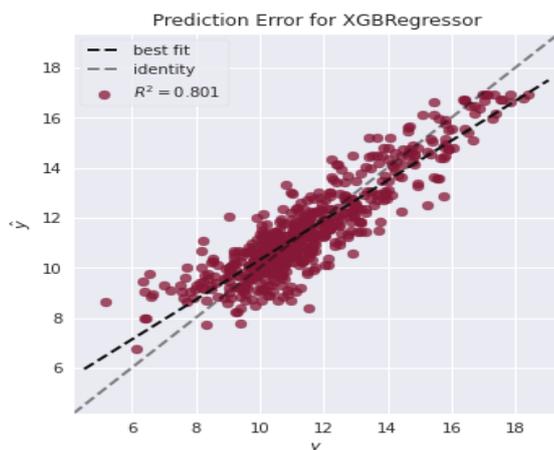


Figure 3 Prediction Error for XGBRegressor

4.4.1.2. Deployment with deep learning framework

Artificial neural network (ANN) represents a large class of machine learning algorithms; their architecture resembles the neurons in the human brain. Keras makes it easy to create models in the Python programming

language through an API and is a higher-level deep learning framework. The goal of Keras is for the creation and training of neural networks, rather than machine learning models in general. The Keras library has many general-purpose functions built in, such as optimizers, activation functions, and layer properties. Deep Networks or Neural Networks generally recommended if the available data size is large. The advantage of neural network is that it has the ability to detect all possible interactions between predictors' variables [28]. It wraps the efficient numerical computation libraries Theano and TensorFlow and allows us to define and train neural network models in a few short lines of code. The same data applied to the previous algorithms by Pycaret applied to the Keras library and gave the good result.

For each training, we used MSE loss function. Our model trained using the optimizer named stochastic gradient descent (SGD) for 60 epochs. Using the backpropagation, the optimizer updates the weight parameters to minimize the loss function. The learning rate is a hyperparameter that controls how much to change the model weights each time. Another powerful regularization technique for neural networks and deep learning models is dropout [29].

4.4.1.3. Definition of the ANN Model

Our model described as bellow:

Type: Regression with Neural Network by Keras

Features: $X = [x_1, x_2]$

Target: $Z = W^T X = \sum_{i=1}^m w_i x_i$ and $W = [w_1, w_2]$

$\hat{Y} = \text{sigmoid}(Z) = 1 / (1 + e^{-Z})$, $Z = W^T X + b$, $b = w_0 x_0$, $x_0 = 1$

Types of layers: Dense Layer

Number of units: (neurons) in each Layer: 2

Number of input layers: 2

Number of hidden layers: 8

Activation Functions: Relu: $g(z) = \max(0, z)$

Loss Function: MSE

Output Type: Value between 0 and 20

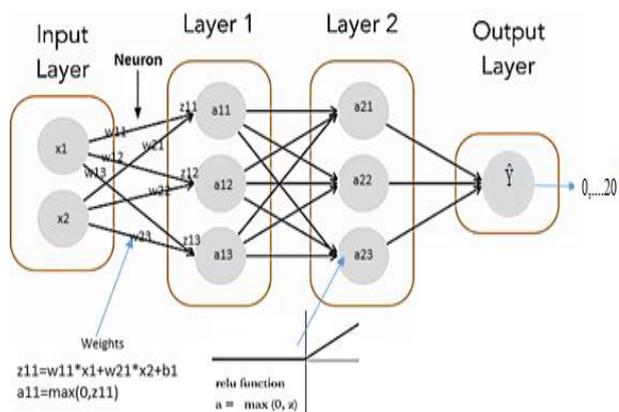


Figure 4 Example of a Neural Network Structure implemented using the Keras.

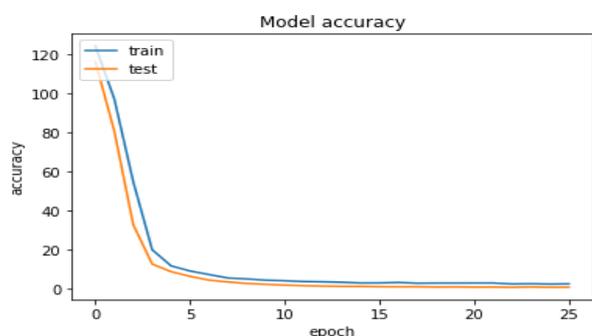


Figure 5 Line plot of Mean Squared Error showing the loss error over the training epochs.

The Figure5 show us that the model converged the tows lines of training and testing. The loss function MSE is a small for a neural network in our model.

4.4.1.4. Implementation

The classroom management, as well as academic and social interactions between students and teachers had a direct influence on students' learning [30]. The teacher is the main pillar for carrying out the task of educational support into their students. Another way to assess the impact of teachers is to calculate the observed averages of the classes taught, in order to study the comparison between them. After extracting, the predictive results from the model, and obtain a list of students who model predict them to fail. The process is not complete without the concerted efforts of the various stakeholders, including managers, inspectors, mentors, and students' parents.

5. CONCLUSION

In this paper, our research mainly focused on predict students' success at the end of their studies using their information's after the first and the second Bacculaureate years as input variables. The development of the model by different approaches such as neural network gives the best

performance as compared to the other existing techniques. The model provides an opportunity to identify aspects of the plan and educational programs that should be improved in order to motivate students to work more seriously and improve their knowledge in specific disciplines. In addition, a teacher can identify and help students who are expected to fail.

The following research can examine the social and health status and history of the educational process to find out the cause of student's dropout and find possible ways to limit or reduce them.

REFERENCES

- [1]. [www.pisa.oecd.org\(2019\)](http://www.pisa.oecd.org(2019))
- [2]. Hellas, Arto, et al. "Predicting academic performance: a systematic literature review." Proceedings companion of the 23rd annual ACM conference on innovation and technology in computer science education. 2018.
- [3]. Ajibade, Samuel & Bahiah, Nor & Shamsuddin, Siti Mariyam.(2019). A Novel Hybrid Approach of AdaboostM2 Algorithm and Differential Evolution for Prediction of Student Performance. International Journal of Scientific & Technology Research.8.65-70.
- [4]. Z. Ibrahim, D. Rusli, Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression, in: 21st Annual SAS Malaysia Forum, 5th September, 2007.
- [5]. [www.men.gov.ma\(2019\)](http://www.men.gov.ma(2019))
- [6]. National Efficiency Assessment Program data PNEA 2016.
- [7]. National Efficiency Assessment Program data PNEA 2016
- [8]. Shah, Saher.(2012). Predicting factors that affect students' academic performance by using data mining techniques.
- [9]. Yasmeen Altujjar, Wejdan Altamimi, Isra Al-Turaiki, Muna Al-Razgan, Predicting Critical Courses Affecting Students Performance: A Case Study, Procedia Computer Science, Volume 82, 2016.
- [10]. Nurafifah, Mohammad Suhaimi, et al. "Review on predicting students' graduation time using machine learning algorithms." International Journal of Modern Education and Computer Science 11.7 (2019).
- [11]. Kasih, Julianti & Ayub, Mewati & Susanto, Sani. (2013). Predicting students' final passing results using the Apriori Algorithm. World Transactions on Engineering and Technology Education.
- [12]. Baradwaj, Brijesh Kumar, and Saurabh Pal. "Mining educational data to analyze students' performance." arXiv preprint arXiv:1201.3417 (2012).
- [13]. Stretch, Pedro; Cruz, Luís; Soares, Carlos; Mendes-Moreira, João; Abreu, Rui, "A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance," International Educational Data Mining Society, p. 4, 2015.
- [14]. ElGamal, A. F. (2013). An educational data mining model for predicting student performance in programming course. International Journal of Computer Applications, 70(17), 22-28.
- [15]. Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers & Education, 61, 133-145.
- [16]. Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. Computers & Education, 68, 458-472.
- [17]. Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 267-270).
- [18]. Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003, November). Predicting student performance: an application of data mining methods with an educational web-based system. In 33rd Annual Frontiers in Education, 2003. FIE2003. (Vol. 1, pp. T2A-13). IEEE.
- [19]. Falakmasir, M. H., & Habibi, J. (2010, June). Using educational data mining methods to study the impact of virtual classroom in e-learning. In Educational Data Mining 2010.
- [20]. González, J. M. B., & DesJardins, S. L. (2002). Artificial neural networks: A new approach to predicting application behavior. Research in Higher Education, 43(2), 235-258.
- [21]. Bauer, J. A., Thomas, T. S., Cauraugh, J. H., Kaminski, T. W., & Hass, C. J. (2001). Impact forces and neck muscle activity in heading by collegiate female soccer players. Journal of sports sciences, 19(3), 171-179.
- [22]. Calvo-Flores, M. Delgado, et al. "Predicting students' marks from Moodle logs using neural network models." Current Developments in Technology-Assisted Education 1.2 (2006):586-590.
- [23]. Prakash, T. Nikil, and A. Aloysius. "Data preprocessing in sentiment analysis using Twitter data." International Educational Applied Research Journal (IEARJ) Volume 3.
- [24]. Yu, Lean, Shouyang Wang, and Kin Keung Lai. "An integrated data preparation scheme for neural network data analysis." IEEE Transactions on Knowledge and Data Engineering 18.2(2005):217-230.

- [25]. Badr, G., Algobail, A., Almutairi, H., & Almutery, M. (2016). Predicting students' performance in university courses: a case study and tool in KSU mathematics department. *Procedia Computer Science*, 82, 80-89.
- [26]. Nathans, Laura L., Frederick L. Oswald, and Kim Nimon. "Interpreting multiple linear regression: A guidebook of variable importance." *Practical Assessment, Research, and Evaluation* 17.1(2012): 9.
- [27]. <https://pypcare.org/> (2019).
- [28]. [Srivastava, et al. in their 2014 paper Dropout: A Simple Way to Prevent Neural Networks from Overfitting], is a technique where randomly selected neurons was ignored during training.
- [29]. Gray, C. McGuinness, P. Owende, An application of classification models to predict learner progression in tertiary education, in: *Advance Computing Conference (IACC), 2014 IEEE International, IEEE, 2014*, pp. 549–554.
- [30]. O'connor, Evelyn A., and Marian C. Fish. "Differences in the Classroom Systems of Expert and Novice Teachers." (1998).