International Journal of Scientific Research in Computer Science and Engineering
Research Paper
Vol.10, Issue.6, pp.22-29, December (2022)
E-ISSN: 2320-7639

Student Grade Prediction by using Machine Learning Methods and Data Analytics Techniques

Anoushka, Shivani Dubey, Vikas Singhal

[1]Dept of Information Technology, Greater Noida Institute of Technology, APJ Abdul Kalam Technical University, Greater Noida, India
[2]Dept. of Information Technology, Greater Noida Institute of Technology, APJ Abdul Kalam Technical University, Greater Noida, India
[3] Dept of Information Technology, Greater Noida Institute of Technology, APJ Abdul Kalam Technical University, Greater Noida, India

Corresponding Author: anoushkatomar30@gmail.com, Tel.: +919870444552

Available online at: www.isroset.org | DOI: https://doi.org/10.26438/ijsrcse/v10i6.2229

Received: 28/Oct/2022, Accepted: 01/Dec/2022, Online: 31/Dec/2022

***Abstract*—** In the world of an open education system, students have the flexibility to learn anything with ease as the learning content is easily available. This can make the student rather confident as well as careless at the same time. Therefore, it becomes challenging to predict the performance of the student beforehand. In this research, an attempt is made to improve the students' situations by predicting their performance in advance. This is done by applying the univariate linear regression model. This would help the students improve their performance based on predicted grades and enable teachers to identify those who need assistance. The Main objective of this paper is to implement a simple algorithmic model that predicts the score an individual student that he/she will get at the end of the year. "G3" or the final grade is our label (output) and the rest of the columns will be our features (inputs).

***Keywords*—** Student Grade System, Data Analytics, Univariate Linear Regression Model

I. INTRODUCTION

In the present world of the open education system, student performance is getting worsen every day gradually. Predicting student performance in the advance can help students as well as their teachers to keep the progress of the student in check. Many educational organizations have adopted continuous evaluation systems for keeping track of their student's performance and hence improving their grades. The main purpose of these continuous evaluation systems is to aid students in their academics. In a continuous evaluation system, tests are conducted on regular basis at a uniform interval of time. In order to have a satisfactory result in the final exam a student should appear in all the tests. The main objective of Student Grade Prediction is to help the students to know their performance in advance by using the un-variance Linear Regression Model [1]. Such techniques would help the students to improve their performance based on the predicted grade and would enable teachers to identify those individuals who might need assistance.

II. RELATED WORK

Universities are acclaimed places for higher education and students' retention in universities is a topic of concern. It has been previously observed that most students tend to drop out of the university during their first year of the course due to a lack of proper support and guidance in undergraduate courses. Due to this reason, the first year of an undergraduate student is referred to as a "make" or "break" year. The lack of any support on the course domain and its complexity tends to demotivate the student. Which in turn may lead to students dropping out of the course. In 2006, it was observed that the yearly school leaving rate in Portugal for 18 to 24 years olds was 40%, while the value in the case of the European Union average was just 15% (Eurostat 2007). The failure in the core subjects i.e. Mathematics and Portuguese (the native language) was more prominent as compared to other subjects like physics and history, as they focus more on the fundamental subjects rather than core subjects. In 2012, after reviewing 13 years of research in order to analyze the factors affecting the Grade Point Average (GPA) of the students, it was observed that the GPA was affected not only by the prior academic performances of the student but also by the demographic factors as well as psychosocial contextual factors. It was also observed that the average GPA of the student mainly depended on self-efficacy which was closely followed by high school GPA, ACT, and grade goal (Psychological Bulletin 2012) [2]. The educational domain offers a vast ground for Machine learning applications since there are numerous sources of data (e.g. old databases, web pages), and disparate interest groups such as faculty, administrators, students, and alumni) [3]. For example, there are various valuable questions that can be answered using the machine learning techniques, such as: Who are the students that are interested in taking more credit hours? What type of courses can be offered to lure more students? What are the main reasons a student transfers? What are the factors that affect student productivity and performance? Is it possible to predict students' grades and performance in advance? What type of student usually returns for more classes? All of these questions can be answered by the use of various machine learning techniques and hence improvise the education system [4]. At present, various types of research have made a point by using various machine learning algorithms as well as DM techniques, for example, Ma et al (2000) applied a DM approach that focused on Association rules for selecting the weak tertiary school of Singapore for remedial classes. The attributes that were selected as input included demographic attributes which included sex and gender and school performances over the past years and a solution was introduced that made the traditional ways of education in Singapore look outdated. In 2003 (Minaei- Bidgoli et al. 2003), all the online grades of the students from Michigan State University were modeled using the three classification approaches (i.e. binary: pass/fail; 3- level: low, middle, high; and 9-level: from 1-lowest grade to 9-highest score). The database consisted of 227 samples with online features and the best results were obtained by a classifier ensemble with an accuracy of 94% (binary) 72% (3-classes) and 62% (9-classes). In 2006, Pardose et al collected data from an online tutoring system regarding USA 8th-grade Math tests [5]. The authors used the regression approach, where the main target was to predict the math test score based on individual skills. They also used the Bayesian networks and the best result was a predictive error of 15% [6]. In this work, we will analyze data collected from two Portuguese schools. The dataset consists of two different types of sources: mark reports and questionnaires. Since there was meager information available under the first topic (i.e. only the grades and number of absences were available), the questionnaire dataset was added to the information, because it had attributes like demographic details, social and school-related attributes (e.g. student's age, father's occupation, parent's education) [7]. The main objective of the project is to predict the final grade of the students and to identify the key factors that affect their academic success/failure rate, from the information available through univariate linear regression [8].

III. METHODOLOGY

According to the Portuguese education system [9], a student should have 9 years of basic education which is followed by 3 years of secondary education which is followed by higher education or degree courses. Similar to many other countries, Portugal's higher education system also uses the 20-point grading system, where 0 is the lowest grade and 20 is the highest grade. During secondary education, the student is evaluated on the basis of three periodic exams or tests, and the last evaluation (G3 of Table 1) is said to be the final grade. This study analyzes the data collected during the year 2005-2006 batch of two Portuguese public schools, from the region of Alentejo. Despite the increase in the usage of information technology in every field, most of the schools in Portugal still relied on pen-paper systems (which was the situation during that time) [10]. Hence the dataset consists of two sources: school reports which included a few attributes (i.e. the three periods grades and number of absences); the questionnaire dataset, which added more information to the dataset. The questionnaire dataset was designed by asking closed questions to the students related to several demographic details (e.g. parent's education, mother's occupation, father's occupation, family income), social and emotional details (e.g. romantic relationship, alcohol consumption) and school-related details (number of past class failures) variables that were expected to affect the student performance. Following this, the questionnaire was reviewed by school professionals and tested on a small group of students which consist of 15 students in order to receive feedback. The resultant questionnaire consists of 37 questions on a single A4 sheet which was answered by 788 students in class. After the process, a total of 111 answers were removed due to the lack of identification details. Finally, the data was integrated into two datasets related to mathematics (395 examples) and the Portuguese language (695 records) classes. At the time preprocessing stage, some of the features were removed due to a lack of discriminative value. For example, a small amount of didn't respond about the family income (probably due to privacy issues) [11]. The remaining attributes are shown in Table 1, where the last four rows denote the variables taken from the school reports. This dataset has been taken into account from the research of P. Cortez and A. Silva in the year 2008.

- **Machine learning models**

Linear regression is a machine learning technique used for modeling the relationship between a dependent and one or more independent variables. When it is applied to single variables it is known as simple linear regression or linear regression. For more than one dependent or explanatory variable, it is known as Multiple linear regression or linear regression with multivariable. Apart from these techniques, Elastic Net regression is a technique used for training the machine learning model which combines the lasso and the ridge regression methods [12]. This machine training technique is the method that learns from shortcomings in order to improve the regularization of statistical models. For the prediction to be more precise and accurate ensemble techniques and randomized decision trees are used. The ensemble technique combines prediction from multiple machine learning algorithms to make a more accurate prediction than a single model. In order to apply this random forest regression, a supervised learning algorithm that used an ensemble learning method for regression is used [13].

Table 1: All the student-related attributes that consist of the dataset.

ATTRIBUTE	DESCRIPTION (DOMAIN)
sex	student's sex (binary: female or male)
age	student's age (numeric: from 15 to 22)
school	student's school (binary: Gabriel Pereira or Mousinho da Silveira)
address	student's home address type (binary: urban or rural)
Status	parent's cohabitation status (binary: living together or apart)
Menu	mother's education (numeric: from 0 to 4[a])
Mjob	mother's job (nominal[b])
Fedu	father's education (numeric: from 0 to 4[a])
Fjob	father's job (nominal[b])
guardian	student's guardian (nominal: mother, father, or other)
famsize	family size (binary: ≤ 3 or > 3)
famrel	quality of family relationships (numeric: from 1 – very bad to 5 – excellent)
reason	reason to choose this school (nominal: close to home, school reputation, course preference, or other)
traveltime	home to school travel time (numeric: 1 – < 15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour or 4 – > 1 hour).
studytime	weekly study time (numeric: 1 – < 2 hours, 2 – 2 to 5 hours, 3 – 5 to 10 hours or 4 – > 10 hours)
failures	number of past class failures (numeric: n if $1 \leq n < 3$, else 4)
schoolsup	extra educational school support (binary: yes or no)
famsup	family educational support (binary: yes or no)
activities	Extra-Curricular activities (binary: yes or no)
paidclass	extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
internet	Internet access at home (binary: yes or no)
nursery	Attended nursery school (binary: yes or no)
higher	Wants to take higher education (binary: yes or no)
romantic	With a romantic relationship (binary: yes or no)
freetime	Free time after school (numeric: from 1- very low to 5- very high)
goout	Going out with friends (numeric: from 1- very low to 5- very high)
Walc	Weekend alcohol consumption(numeric: from 1- very low to 5- very high)
Dalc	Workday alcohol consumption (numeric: from 1- very low to 5- very high)
health	Current health status (numeric: from 1- very bad to 5- very good)
absences	Number of school absences (numeric: from 0 to 20)
G1	First period grade (numeric: from 0 to 20)
G2	Second period grade (numeric: from 0 to 20)
G3	Final grade (numeric: from0 to 20)

Apart from all these algorithms, SVM or support vector machines are supervised learning models with associate learning algorithms that analyze data used for classification and regression analysis. Following all of these, in order to reduce the errors in the training models, the gradient-boosted algorithm is used. This algorithm reduces the error in subsequent models, by taking errors in the previous models into consideration [14]. The flow that is followed in analyzing the data and making predictions is mentioned in fig.1.

1. Reading the Dataset: The first step is to read the dataset from the client or user. In this case, the dataset is the students' data that was collected from the school which can be either in CSV format or an Excel sheet. In this case, it is a CSV file that is read with help of the pandas module.

2. Finding the dependency of the mentioned attributes on the final grade attribute (G3): The second step is to find the correlation between different variables or attributes mentioned in the dataset with the final grade (G3). The correlation values range from -1 to 1. If the value is negative, the attribute is inversely related to variable G3 and if the value is positive, it is directly related to the variable G3. If the value is an extreme value or the domain value, then it is highly correlated with the G3 attribute.

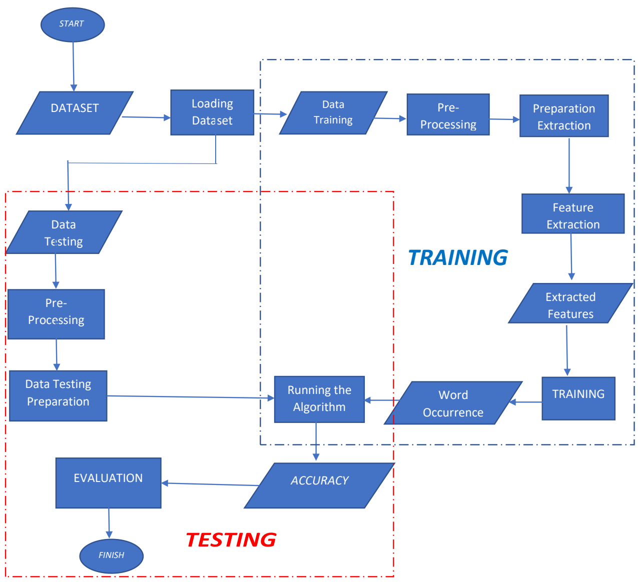

Fig 1. Flowchart for the whole processing

3. Irradicating the least correlated attributes: The third step is to remove the attributes that are least correlated with the attribute G3. This is step is important in order to achieve more accuracy. For better computational results the least correlated attributes are removed from the file.

4. Splitting the data for training and testing: This step is one of the most important and crucial parts of the project. Here we import the sklearn module for training and testing the data as well as performing our prediction algorithm. We perform the training and test the data by sending X and Y labels. The X and Y labels are the parameters on which the training and testing of the data are computed.

5. Prediction: After the data is trained and tested, now it is fitted using the Linear regression algorithm. They fitted using the Linear regression algorithm. The fit() function is used to set the accuracy of the computed values. The fit() function gives the desired results/values. The predict() is used for prediction which too uses the Linear regression model.

6. Graphical Representation of the Result: The final step is the graphical representation of the predicted results. The graph which is being used is the “box-plot” graph. Here, we are comparing the given final grade with the predicted results which is also the final grade (G3). The reason for using boxplot is that it shows the representation accurately in terms of statistics i.e. mean, median, quartiles, etc. through this we can say if the predicted values are accurate or not.

IV. RESULTS AND DISCUSSION

After irradicating the least correlated values with the attribute G3, it was important to ensure that there are no null values left in the dataset before starting the processing. In order to achieve that a heatmap was plotted. The next step in data visualization was to see if the gender ratio was equal or had a difference. So a count plot graph was plotted as shown in fig 2. As a result, it turned out that the gender distribution is pretty much even. The next step was considering age as a factor affecting the grades of the student. so as a first step, the kernel density estimation was done in order to see the number of students belonging to different age groups starting from the age of 14-22 as shown in fig 3. After that, the age groups were further divided into two parts i.e. male and female in order to find the gender ratio in different age groups as shown in fig 4 [15]. After plotting the count plot of that, it turned out that the gender distribution between males and females was pretty even between the age of 15-19 but for the age group above 19 years old, the gender distribution was low and unequal and the conclusion was made that the students above 19 years old may be outliners, year back students or dropouts. The next part that was considered for processing was finding out if the grades of the students depend upon their demographic settings i.e. if the student lives in a rural or urban area. In order to do that, a count plot graph was plotted to see the ratio of students living in urban and rural areas as shown in fig 5. As a result, it turned out that approximately 77.72% of students came from urban. areas and the remaining 22.28% came from rural areas [16]. The next step in order to predict if the demographic setting of a student affected the final grade, a count plot graph was plotted to see how many students were able to achieve a high score that is living in an urban setting compared to the students that are living in a rural setting as shown in fig 6. As a result, it turned out that the students living in an urban setting were achieving higher scores as compared to the students living in a rural setting [17].

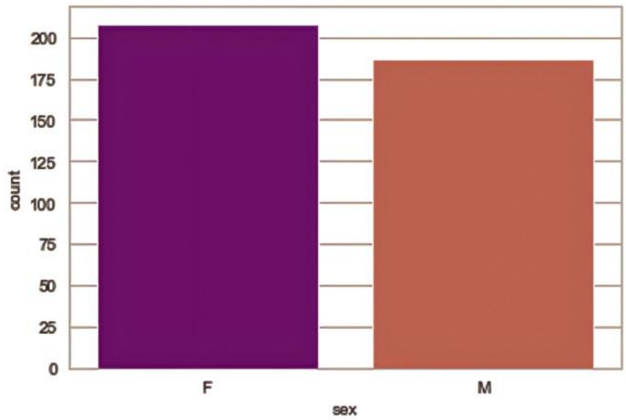

Fig 2. The gender distribution count plot graph

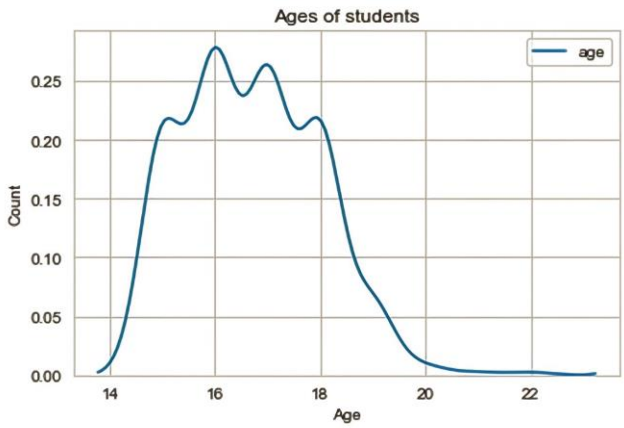

Fig 3. The age group division graph

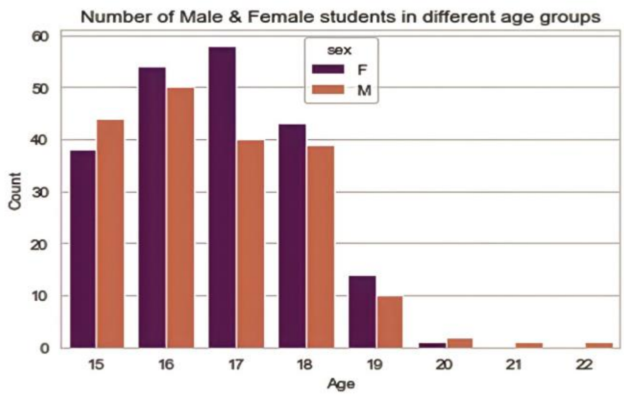

Fig 4. Count plot graph dividing male and female students into the different age groups

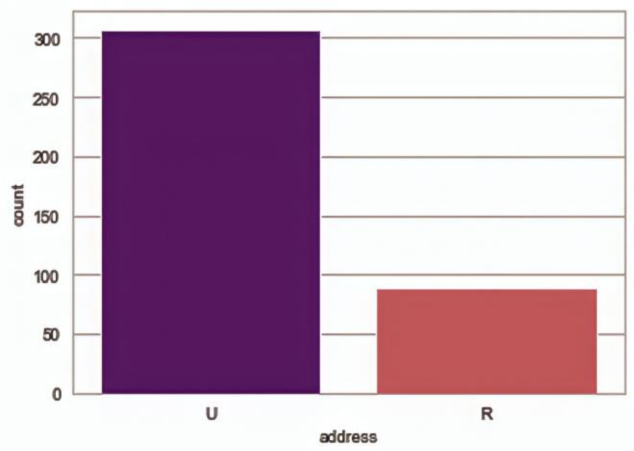

Fig 5. Count plot graph for dividing students living in urban and rural areas

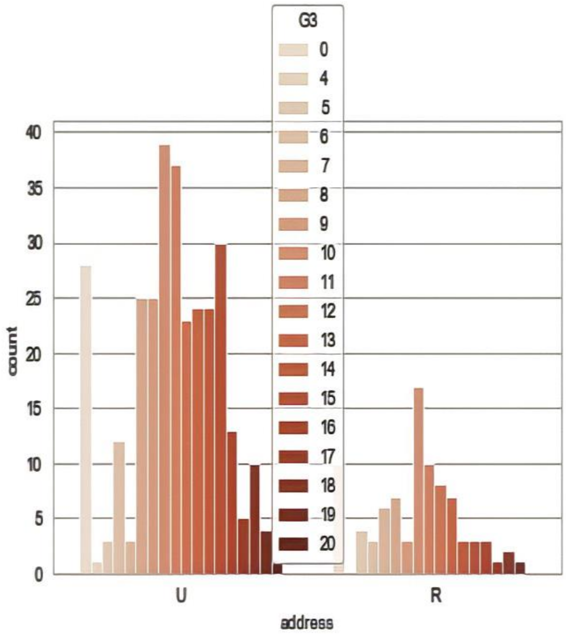

Fig 6. Count plot graph of address vs G3 grade

After all the pre-processing was done, the exploratory data analysis was done. The first step in performing the exploratory data analysis, is to find out, “Does age affect the final grade?”. In order to find the answer to this question, a boxplot graph is plotted as shown in fig 7. The boxplot graph shows that the median grades of the three age groups (15, 16. 17) are similar. It is noted that the skewness of age group 19 and age group 20 seems to score higher grades among all. The step in performing the exploratory data analysis is to find the answer to the question “Do urban students perform better than rural students?. For answering this question, a kernel density estimation graph is plotted as shown in fig 8. As a result, it clearly shows that there is not much difference between the grades based on location i.e. the students staying in urban areas approximately score the same score as students residing in rural areas [18].

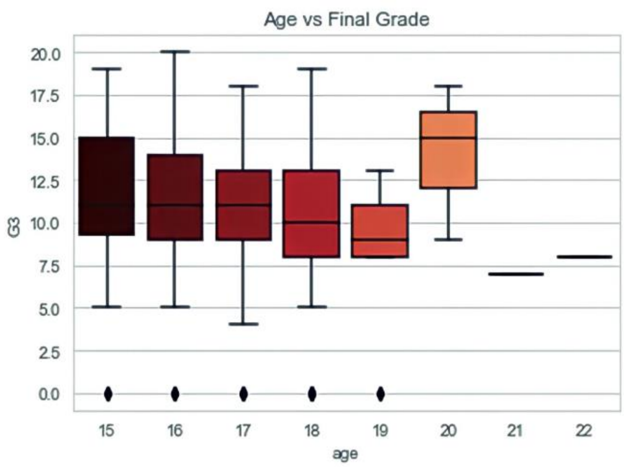

Fig 7. The Age vs Final Grade box plot Graph

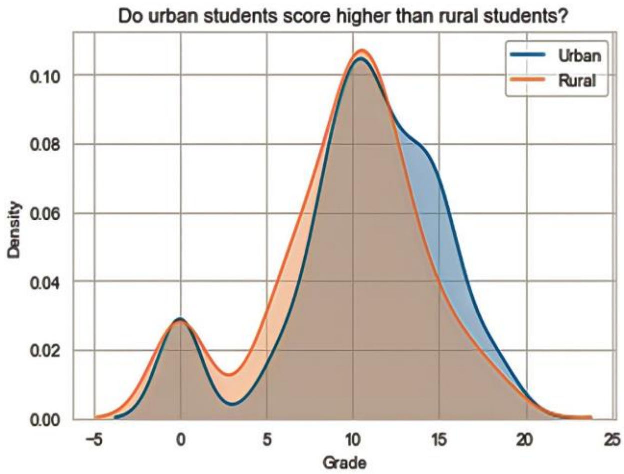

Fig 8. The Urban vs Rural area students score comparison graph.

After the exploratory data analysis is done, the next step is encoding categorical variables using LableEncoder(). First of all, we drop the variables that are highly correlated with the grade G3 i.e. the variables G1 and G2, the period grades of the student. Although is much more difficult to predict G3 without G2 and G1, such prediction is much more useful because our goal is to find other factors that affect the final grade. The first variable we will consider for prediction is the failure attribute. In order to find if the past failures affect the final grade we plot a swarm plot graph targeting both variables i.e. G3 and previous failures as shown in fig 9. It is observed that the student with less number of previous failures usually scores higher as compared to the students with more past failures. The next attributes that are considered for the prediction are the Family education attribute which consists of the variables of the Father’s education and the mother’s education. In order to find if family education affects the final grade a swarm plot graph is plotted as shown in fig 10. It is observed that students that belong to educated families tend to score higher grades [19]. In order to achieve that a box plot graph is plotted as shown in fig 11. It is observed that the students who wish to go for higher studies score more than the students who don’t have a goal in mind. The next attribute that is taken into consideration is the “going out with friends" attribute. For doing this a swarm plot graph is plotted after analyzing the number of students that as shown in fig 13. It is observed that the student that is frequently going out scores fewer marks as compared to the students who choose to stay in and study. The next attribute that is considered is the romantic relationship attribute. For predicting if a romantic relationship affects the final score, a swarm plot graph is plotted as shown in fig 14 [20]. In the graph, the romantic attribute with a value of 0 means no relationship, and a value of 1 means no relationship. It is observed the students with no romantic relationship score higher as compared to the students that are not in a romantic relationship. The last attribute that is considered in order to predict the final score is the "reason" attribute i.e. the students who choose the school with a reason to the students who didn't have any reason to choose the school. In order to predict that, a swarm plot graph is plotted as shown in fig 15. After plotting, it is observed that the students have an equally distributed average score when it comes to the reason attribute i.e. the reason attribute does not affect the final grade [21].

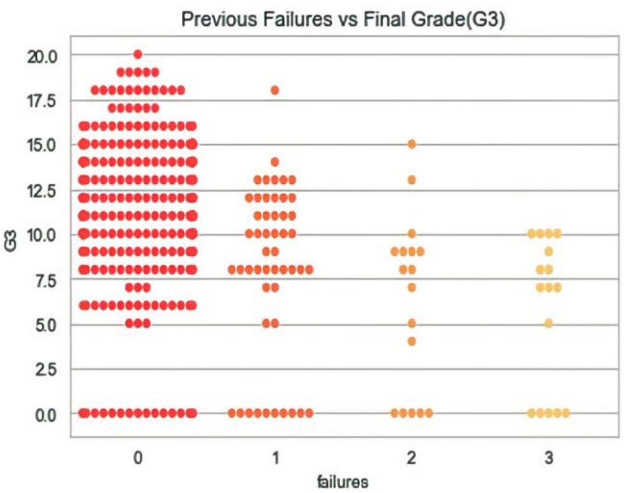

Fig. 9. Personal failures vs Final Grade Swarm Plot Graph

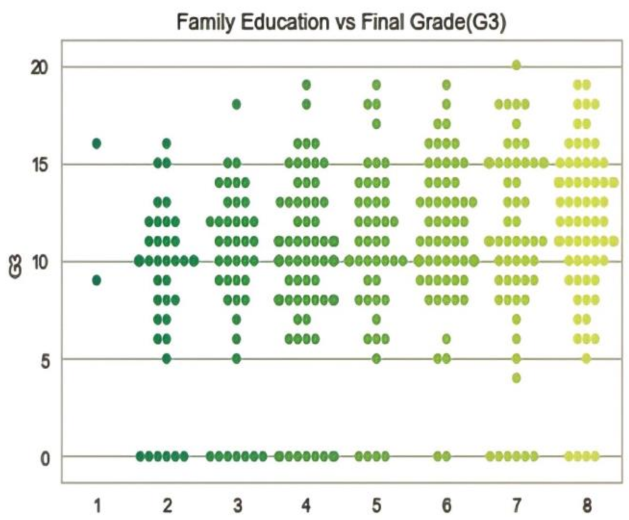

Fig 10. Family education vs Final Grade Swarm Plot Graph

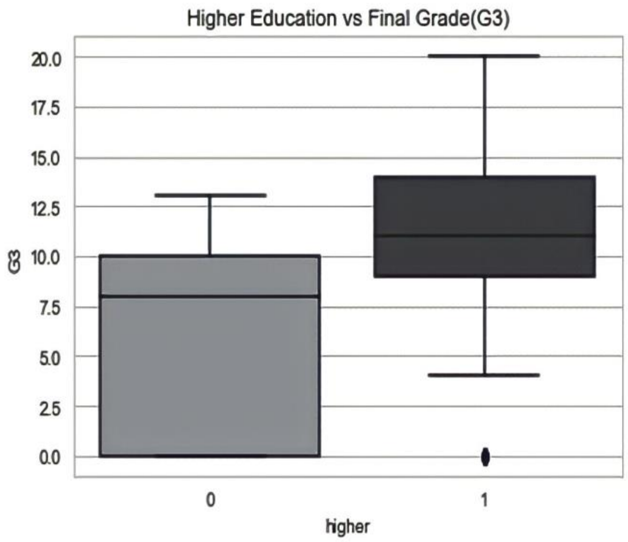

Fig 11. Higher Education vs Final Grade Box plot Graph

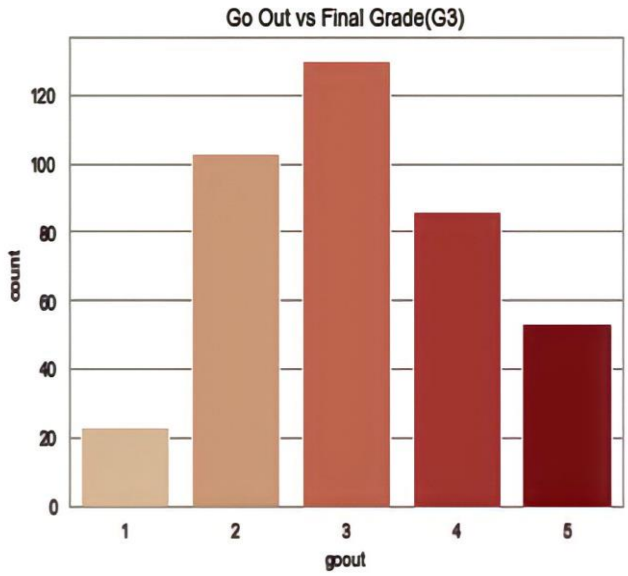

Fig 12. Go out vs Final Grade count plot graph

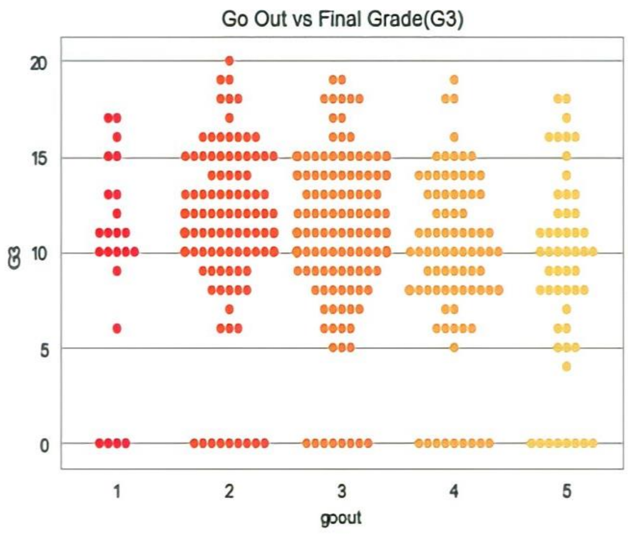

Fig 13. Go out vs Final Grade swarm plot graph

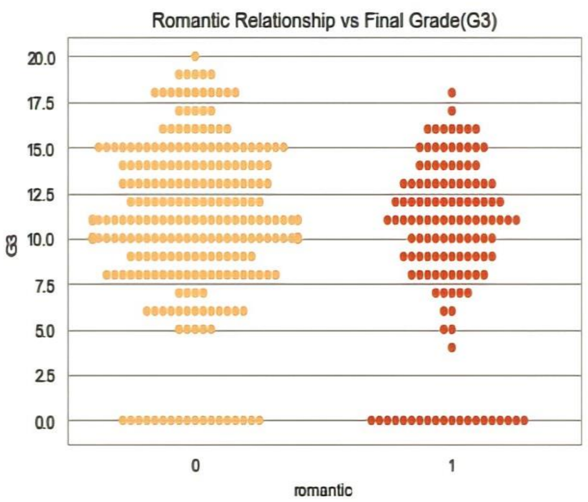

Fig 14. Romantic Relationship vs Final Grade Swarm Plot Graph

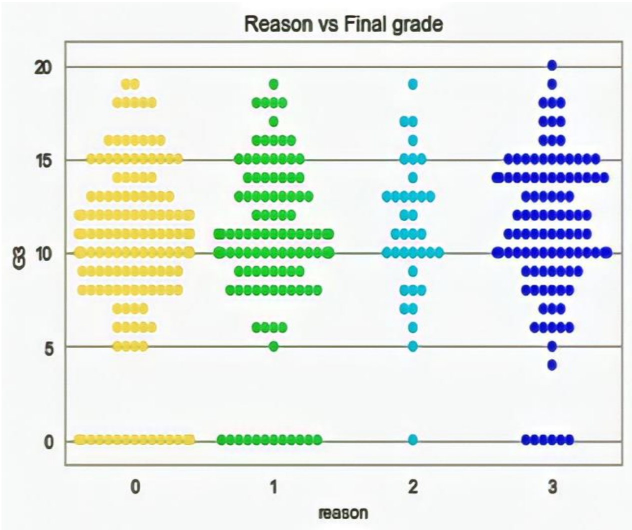

Fig 15. Reason vs Final Grade Swarm Plot Graph

After making all the predictions and observations, the next step is to calculate the mean absolute error and root mean square error of various machine learning algorithms which include linear regression, Elastic Net regression, Random forest, Extra trees, SVM, Gradient Boosted, and Baseline as shown in table 2. After that, the results from the table are plotted using a subplot graph as shown in fig 16. The observation that is made by seeing both the table and the graph is that the linear regression is performing the best in both cases [22].

Table 2. Results of evaluation of different machine learning algorithms

Methods	mae	rmse
Linear Regression	3.48512	4.4326
ElasticNet Regression	3.60805	4.57327
Random Forest	3.72601	4.61621
Extra Trees	3.7797	4.77882
SVM	3.54927	4.58147
Gradient Boosted	3.57244	4.50059
Baseline	3.78788	4.82523

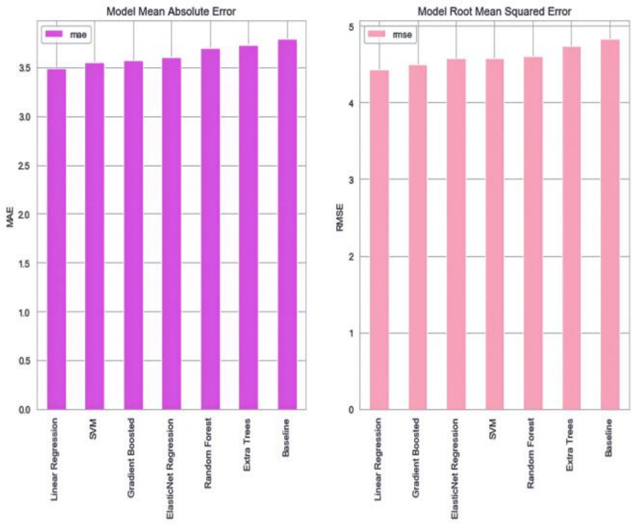

Fig 16. Model Mean Absolute Error and Model Root Mean Squared Error of different machine learning algorithm

V. CONCLUSION AND FUTURE SCOPE

Education is a very important element in order to grow in society. Using machine learning methods and data visualization methods in daily life, can strengthen the education field and can help the weaker students too in finding the path which is best suitable for them. Machine learning methods can improve the quality of education and can also help in improving resource management. In this paper, we have addressed the prediction of the final grade of the students of two Portuguese schools, by using the past grades, demographic details, and social and other school-related data. Many machine learning algorithms, such as Linear Regression, Elastic Net Regression, Random Forest, Extra Trees, SVM, Gradient Boosted, and Baseline were explored. Also, distinct input selections (with or without the past grades) were also explored in order to obtain more precise and correct predictions. The obtained results reveal various conclusions based on the variable that is studied corresponding to the attribute G3. According to the observation, it is seen that the student's grades are highly affected depending upon the past failures attribute i.e. the more number of past failures, the more the chances of a low final score. The next observation that is being made is that the grades of students depend upon the family education i.e. father's education as well as the mother's education. The students belonging to highly educated family background score higher grades. It has also been observed that the student's ambition for higher studies also affects the final grade of the student i.e. the students that wish to go for higher studies after high school tends to score higher scores. This confirms the conclusion found by P. Cortez and A. Silva 2008.

ACKNOWLEDGMENT

I would like to thank, my college guide and my project coordinator who helped me in regulating my study by understanding my project and helping in collecting more data and improving my study. Apart from this, I would like to give my sincere thanks to my parents who always support me in achieving all my goals.

REFERENCES

[1] P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) EUROSIS, ISBN 978-9077381-39-7. Porto, Portugal, pp.**5-12**, April **2008**.

[2] W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, pp.**123–135, 1993.**

[3] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., Oxford: Clarendon, Vol.2, pp.**68–73, 1892.**

[4] Kotsiantis S.; Pierrakeas C.; and Pintelas P., 2004. Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. Applied Artificial Intelligence (AAI), 18, no. 5, 411–426B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished, **2004**.

[5] Pritchard M. and Wilson S., 2003. Using Emotional and Social Factors To Predict Student Success. Journal of College Student Development, 44 no 1, **18-28, 2003.**

[6] Merceron, A. and Yacef, K., 2010. Measuring correlation of strong symmetric association rules in educational data. C. Romero, S. Ventura, M. Pechenizkiy, & RSJ d. Baker (Eds.), Handbook of educational data mining, pp.**245-256, 2010.**

[7] Heemskerk, I., ten Dam, G., Volman, M. and Admiraal, W., 2009. Gender inclusiveness in educational technology and learning experiences of girls and boys. Journal of Research on Technology in Education, Vol.**41**, Issue.**3**, pp.**253-276, 2009.**

[8] Rohit Raja, Tilendra Shishir Sinha, Raj Kumar Patra and Shrikant Tiwari(2018), Physiological Trait Based Biometrical Authentication of Human-Face Using LGXP and ANN Techniques, Int. J. of Information and Computer Security, (Scopus Index), Vol.**10**, Nos. 2/3, pp.**303- 320, 2018.**

[9] Rohit Raja, Tilendra Shishir Sinha, Ravi Prakash Dubey (2015), An Empirical Analysis for Detection of Occlusion for face image parallel to the surface plain using Soft-Computing technique, Mats Journal of Engineering & Vol.I (1), pp. 1-6Technology, ISSN 2394-0549. Vol.**1**, Issue.**2**, pp.**95-102, 2015.**

[10] M. Mayilvaganan, D. Kalpanadevi, Comparison of classification techniques for predicting the performance of students academic environment, in: Communication and Network Technologies (ICCNT), International Conference on, IEEE, pp.**113–118, 2014.**

[11]]K. F. Li, D. Rusk, F. Song, Predicting student academic performance, in: Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on, IEEE, pp.**27–33, 2013.**

[12] M. Mayilvaganan, D. Kalpanadevi, Comparison of classification techniques for predicting the performance of students academic environment, in: Communication and Network Technologies (ICCNT), 2014 International Conference on, IEEE, pp.**113–118, 2014.**

[13] W. Ham¨ al¨ ainen, ¨ M. Vinni, Comparison of machine learning methods for intelligent tutoring systems, in: Intelligent Tutoring Systems, Springer, pp.**525–534, 2006.**

[14] 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**.

[15] T. Wang, A. Mitrovic, Using neural networks to predict student's performance, in: Computers in Education, 2002. Proceedings. International Conference on, IEEE, pp. **969–973, 2002.**

[16] M. Mayilvaganan, D. Kalpanadevi, Comparison of classification techniques for predicting the performance of students academic environment, in: Communication and Network Technologies (ICCNT), International Conference on, IEEE, pp.**113–118, 2014.**

[17] S. T. Jishan, R. I. Rashu, N. Haque, R. M. Rahman, Improving accuracy of students final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique, Decision Analytics 2 (1), pp.**1–25, 2015.**

[18] C. Romero, M.-I. Lopez, ´ J.-M. Luna, S. Ventura, Predicting students' final performance from participation in on-line discussion forums, Computers & Education 68, pp.**458–472, 2013.**

[19] Bindushree V., Rashmi G.R., Uma H.R., "Analysis of Text Recognition with Data Mining Techniques," *International Journal of Scientific Research in Computer Science and Engineering*, Vol.**7**, Issue.**6**, pp.**40-42, 2019.**

[20] Manimaran R. and Vanitha M, "An Efficient Study on Usage of Data Mining Techniques for Predicting Diabetes", International Journal of Advanced Research Trends in Engineering and Technology (IJARTET), ISSN: 2394-3785, Vol.**3**, Issue.**20**, **pp.268-272, 2016.**

AUTHORS PROFILE

Ms. Anoushka has completed her secondary education from Mount Olivet Sr. Sec. School, Delhi in 2019. She is currently pursuing B.Tech in Information Technology from Greater Noida Institute of Technology, Greater Noida, Uttar Pradesh which is affiliated with APJ Abdul Kalam Technical University. She has worked on various projects that go from developing a morse code application in python to a blogging website in her four years of the Engineering program. She has also developed many applications that range from a movie app to a simple quiz application in order to find her specific domain for doing research.

Dr. Shivani Dubey, Associate Professor in the Department of Information Technology in GNIOT, Greater Noida has more than 16 years of experience in academics and teaching. She has done her doctorate from Ansal University, Gurgaon. Her specialization is in cloud computing, data science, distributed systems, and supply chains. She has published 42 research papers in reputed conferences and journals.

Prof. Vikas Singhal, Head of the Department of Information Technology in GNIOT, Greater Noida. His main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT, and Computational Intelligence based education. He has more than 22 years of teaching experience and 4 years of Research Experience.

