Predict high school students' final grades using basic machine learning.

Sigit Sugiyanto [1,*]

[1] Universitas Muhammadiyah Purwokerto, Indonesia
[1] sigitsugiyanto@ump.ac.id
* corresponding author

(Received December 21, 2020, Revised January 5, 2021 Accepted January 14, 2021, Available online January 15, 2021)

Abstract

To improve the quality of students, teachers must be able to take precautionary measures to deal with students who are lacking or have the potential to experience deficiency. Student ratings are temporary, however, have a profound impact on students' mental and enthusiasm for learning. As a teacher, it is very important to make predictions in dealing with this matter because if the ranking has been issued, it is too late. By using MAE (Mean Absolute Error), this study got a value of 1.09 which is very close to the value in the test, it showed how close our prediction models were with a very small percentage of error.In this article, we will discuss and make Student grade predictions using basic machine learning, we will also discuss the continuity between student data and machine learning. We hope this research will help teachers, people in the education industry or even parents to know more about predicting student final grades and can help them take preventive action.

Keywords: MAE, Machine Learning, Student Grade Prediction, Educational Data Mining

1. Introduction

As universities are prestigious places for higher education, student retention in these universities is of high concern to both students, parents and teachers. A significant number of students have been found to drop out of university during their first year because of a lack of basic support for undergraduate programs. For this reason, a "make or break" year is labeled for the first year of undergraduate students. It can demotivate the student without having some assistance on the course area and its difficulty and can be a reason for canceling the course. There is a tremendous need to establish suitable strategies in higher education institutions to support student retention. Primary grade prediction is one solution that tends to track student success in university degree lectures that can lead to an improvement in student performance based on expected grades.

It will boost student performance by using machine learning with Educational Data Mining (EDM). In enrolled courses, various models may be built to forecast student grades, and provide useful evidence to promote student retention in such courses. Based on what the system will do to encourage the teacher to pay particular attention to these students, this knowledge can be used to classify students at risk early on. In order to better monitor their performance in a way that can improve student retention rates in universities, this knowledge can also support in predicting student grades across different subjects.

2. Literature Review

2.1. Educational Data Mining (EDM)

Stated by [1][2], Educational Data Mining (EDM) focuses on data mining being implemented to academic data. In most situations, EDM is close to standard data mining. We can however, take into account the unique features of academic datasets. The data also has several layers. Many experiments are often carried out at a single institution, which does not make it very straightforward to generalize most studies [3-4].

2.2. Classification

By learning a decision boundary in a dataset, a classification task assigns a category/class to each sample [5]. This datasets is called a datasets of training and includes samples for each subject and the desired class/category [6,8]. The training datasets provide "features" as columns, and a mapping is learned for each sample between these features and the target. Using a test datasets to assess the efficiency of mapping (which is separate from the training datasets) [7]. Test datasets contain only the columns for the functions and not the objective columns. Using the mapping gained on the training datasets, the objective column is projected. In this example, using a training dataset, we will use a classifier to train a model, forecast the goals for the test datasets, and simulate the results using plots[8-9].

2.3. Regression

The targets are integers for classification. Even so, when the targets in a dataset are actual values, regression becomes the function of machine learning. Any sample has a real-valued output or goal in the datasets [10]. This illustrates how a curve (regression) is fitted, which represents much of the data points (blue balls). The curve is a straight line here (red). The role of regression is to consider this curve that defines the underlying distribution of data points [11]. The target for a new sample will lie on the curve learned by the regression task. A regressor finds the mapping between the features of a datasets row and its target value. Inherently, it aims to match the targets with a curve.

2.4. Linear Regression Model

By fitting a linear equation to observed data, linear regression attempts to predict the relationship between two variables. An explanatory variable is assumed to be one variable, and a dependent variable is considered to be the other. For example, using a linear regression model, a modeler will choose to link the weights of individuals to their heights [12].

The researcher must first determine whether there is a relationship between the variables of interest before trying to adjust the linear model to the observed data [13]. This does not necessarily imply that the other is caused by one variable (As example higher SAT scores didn't give students a higher grade on colleges), but that the two variables have some significant association. While evaluating the relation between two variables, a scatterplot can be a useful tool. When there appears to be no connection between both the explanatory and dependent variables proposed, i.e. no increasing or decreasing trends are shown in the scatter plot, then it is likely that a helpful model is not provided by fitting a linear regression model to the data[14]. A valuable numerical measure of the association between the two variables is the correlation coefficient, which is a value between -1 and 1 indicating the strength of the association of the data observed for the two variables[15]. A linear regression line has an equation of the form $\boldsymbol{Y = a + bX}$, where $\boldsymbol{X}$ is the explanatory variable and $\boldsymbol{Y}$ is the dependent variable. The slope of the line is $\boldsymbol{b}$, and $\boldsymbol{a}$ is the intercept (the value of $\boldsymbol{y}$ when $\boldsymbol{x} = 0$).

3. Method

In this study, we're going to start by finding some relation and important data that can be used to create the model for grade prediction later. First, we're going to introduce you to the data set information that we use in this paper, this data set is achieved from the UCL Machine learning repository, Paulo Cortez from the University of Minho is the one who submitted this data set. This data approaches student achievement at two Portuguese schools in secondary education. eStudent grades, demographic, social, and school-related characteristics are included in the data attributes and we gathered it using school records and questionnaires. Two datasets on performance in two different subjects are provided: mathematics (mat) and Portuguese (por). Under binary/five-level classification and regression tasks, the two datasets were modeled. Interesting mention: the G3 target attribute is closely correlated to the G2 and G1 attributes. This is because G3 is the grade of the final year (issued in the 3rd period), whereas G1 and G2 are the grades of the 1st and 2nd years.

ISSN 2723-6471

Without G2 and G1, it is harder to forecast G3, but such a prediction is far more useful. We are going to do some student grade analysis to start this research.

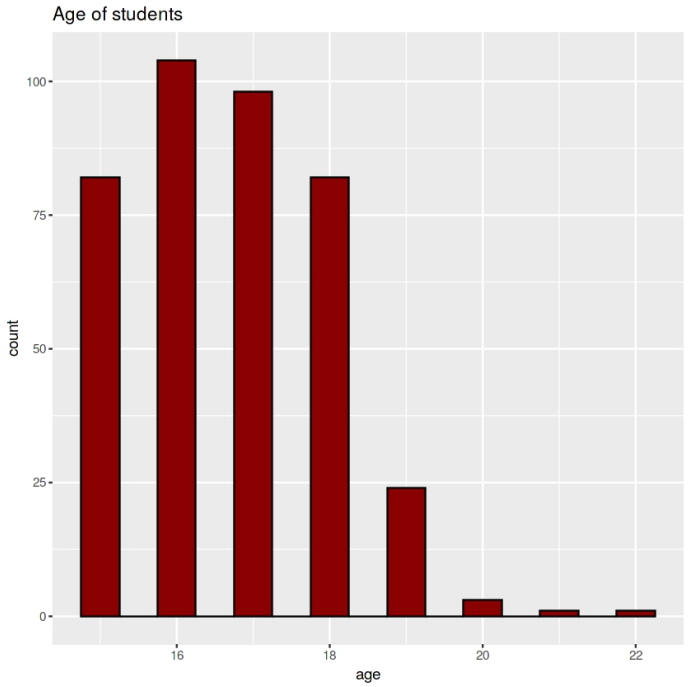

Figure. 1. Age of students

From the fig. 1 above, we can see that most students on this dataset are between age 15 and 18, which quite makes sense since usually most students start their high school at age 15 and eventually graduate by age 18 given the fact that every high school around the world only lasts 3 years. However, in this data set, 29 students are older than 18 years old. and, it would be very interesting to identify their gender. But, before that, we need to explore more about the gender differences in this dataset concerning G1 (First-period grade). What we are going to do is:

- Checking the differences between a female and male student in the school
- Analysis of the class performance based on their gender and age with answering this question below:

a) Between female and male students, who perform better?
b) Did student age affect their performance?

ISSN 2723-6471

3.1. Checking gender differences

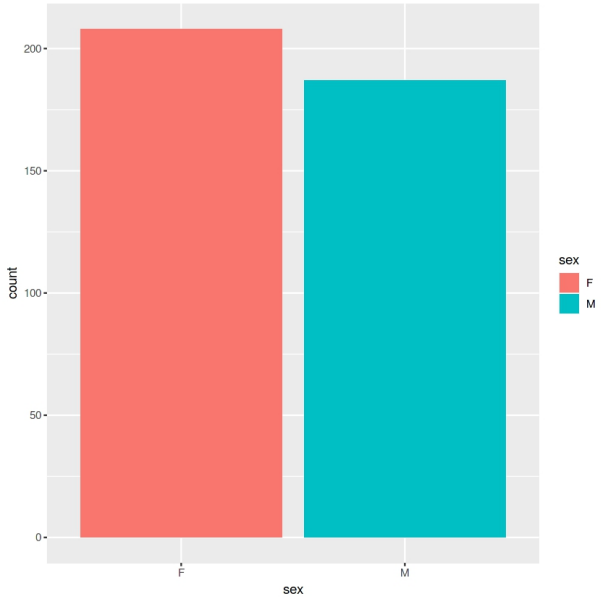

Figure. 2. Gender differences

From the figure above we can see that the female student is a bit dominant rather than the male student, there are 208 female students and 187 male students. Next, what about student health conditions? Figure 3 shows that most of the students are in good health. And, where did they live? There is no specified location about where they live but we can provide some data that indicates that most of the students live in the urban area (See figure 4).

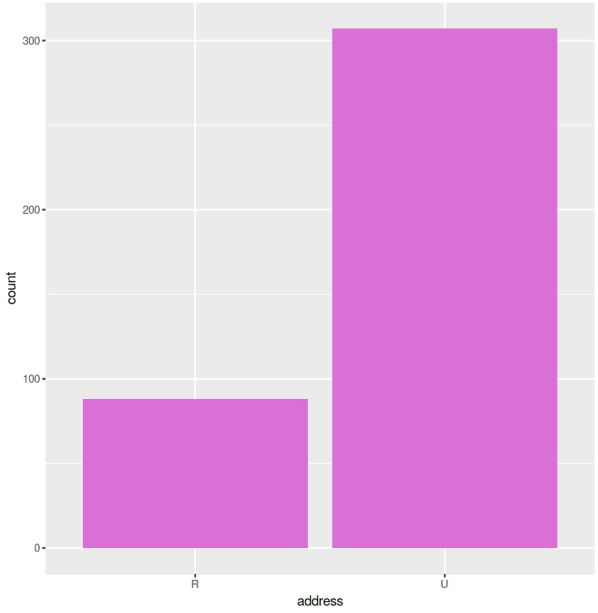

Figure. 3. Student health

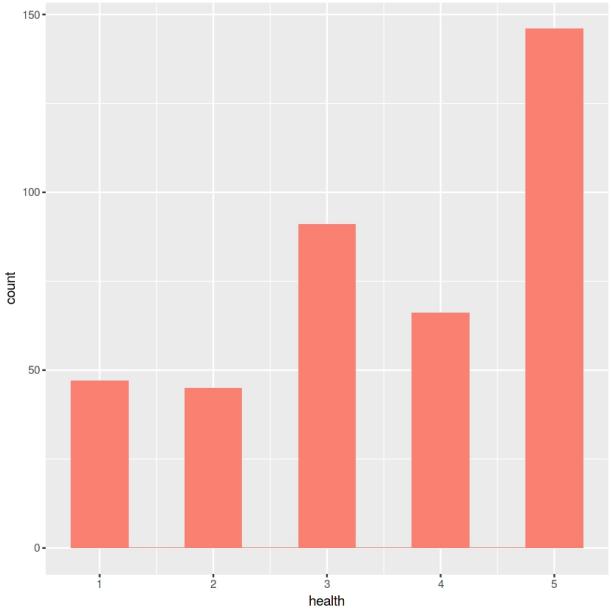

Figure. 4. Student Address

As we can see in figure 5, it is interesting that it's shown in the graph, most of the students that are older than 18 from figure 2 before are male students, which means there are no female students older than 20 years old.

ISSN 2723-6471

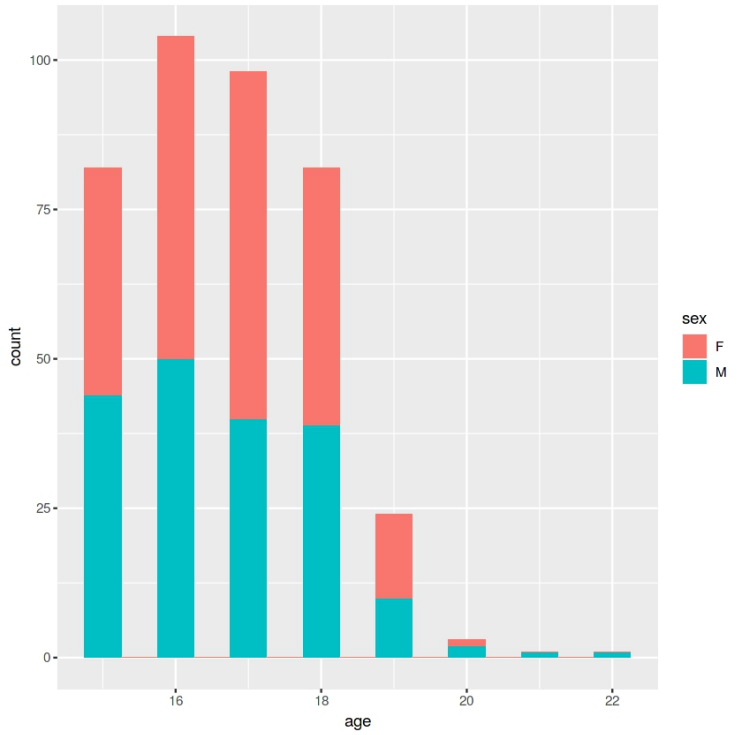

Figure. 5. Gender Age

3.2. Class performance analysis

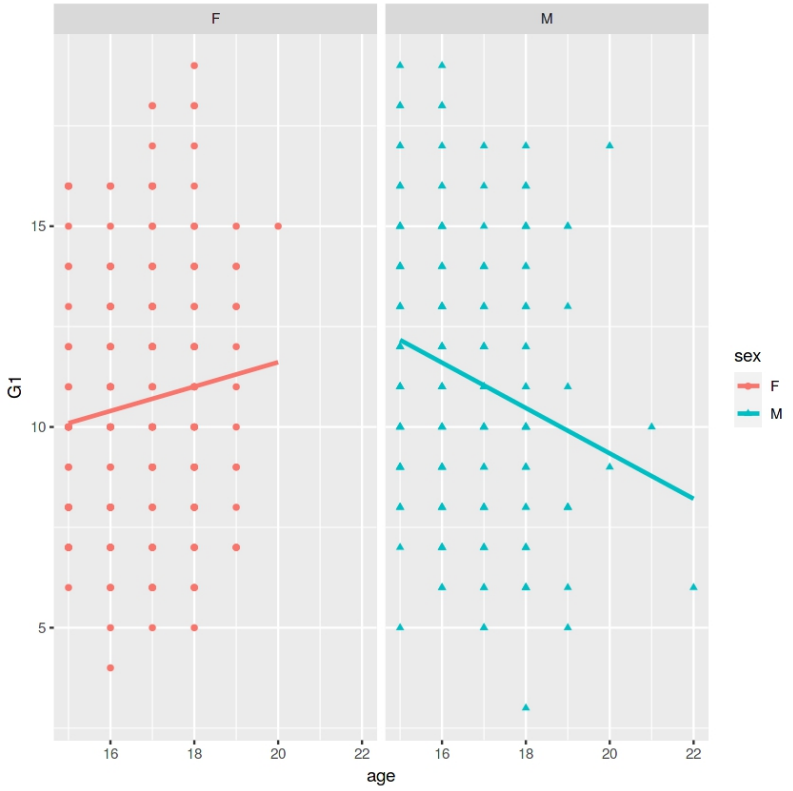

Figure. 6. Gender Performance

From the graph above (Figure 6) we can see that Female students' performance gets better with age, however, we noticed that there is a decrease in male students' performance. The American psychological association's

ISSN 2723-6471

meta-analysis investigates gender gaps in classrooms across 300 countries and over almost a half century. And, their research and analysis proved that female students have been getting better performance than male students over the decades. Female students not only perform better in language class but also outperform male students in math and sciences class. The next question that comes into our head is Why do male students' performance decrease with age? and then that question also give us some suggestion like:

a) Does attendance matter? or is missing or absence during class can negatively affect a student's academic performance?
b) Does giving students too many holidays have any impact on student performance?
c) Do students who live near the school have better results than those who live far from schools?

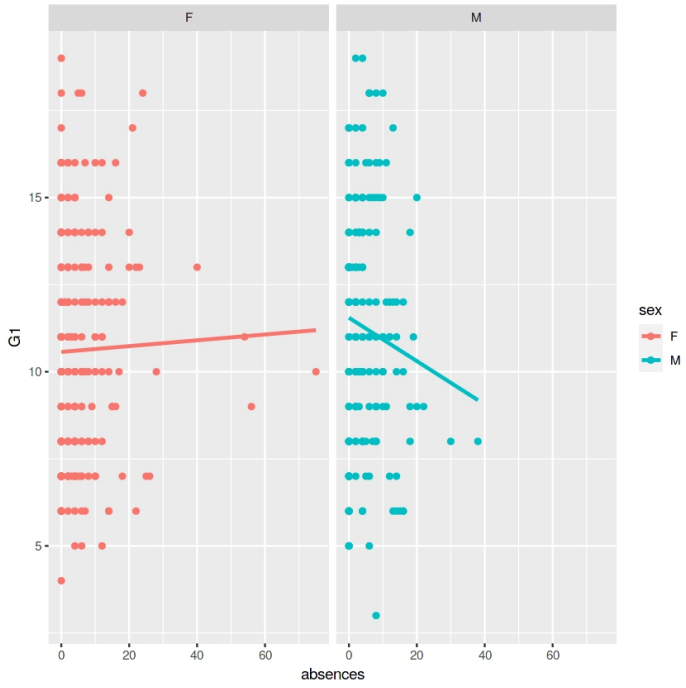

Figure. 7. Absence effect

The graph above (Figure 7), shows that absence can negatively affect male student performance. However, missing any class doesn't have any negative impact on female student performance in class. And then how about holidays affect their performance? As shown in the graph below, holidays also negatively affect male student performance, the further the male student resides, the less result he gets. We have been analyzing holidays and absence, and from the result we got, we can identify them as a major factor affecting the male students' grades.

ISSN 2723-6471

What about parenting behavior and educational support that can affect student performance? to answer that question we gonna separate that question into a different piece like:

a) Do a parent's educational level and a job can affect their child's performance in school?

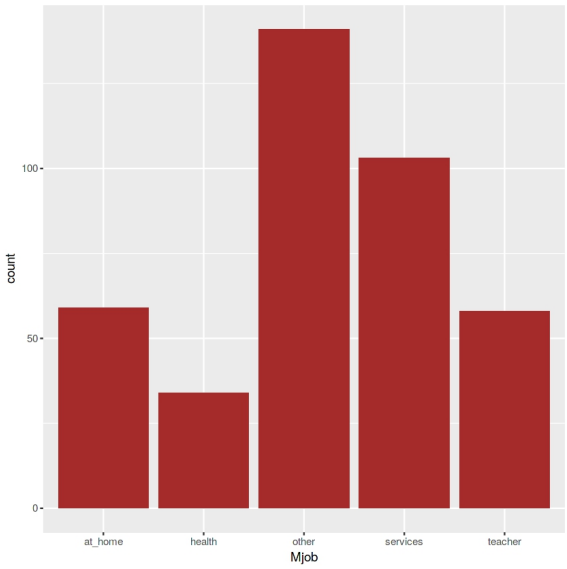

Figure. 8. Mother job

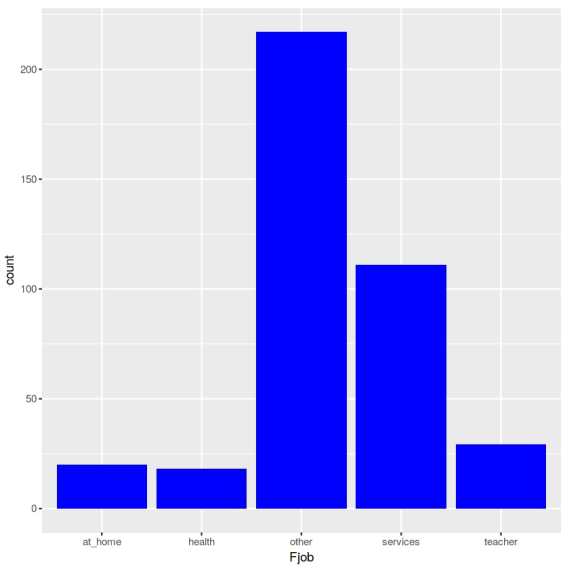

Figure. 9. Father Job

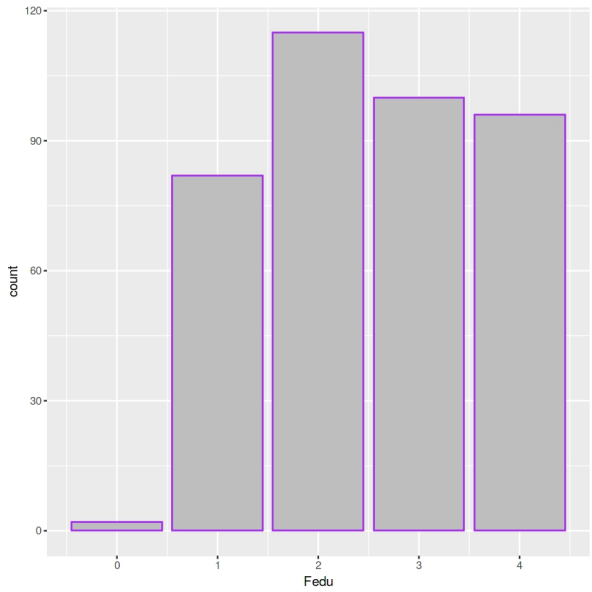

Figure. 10. Mother education

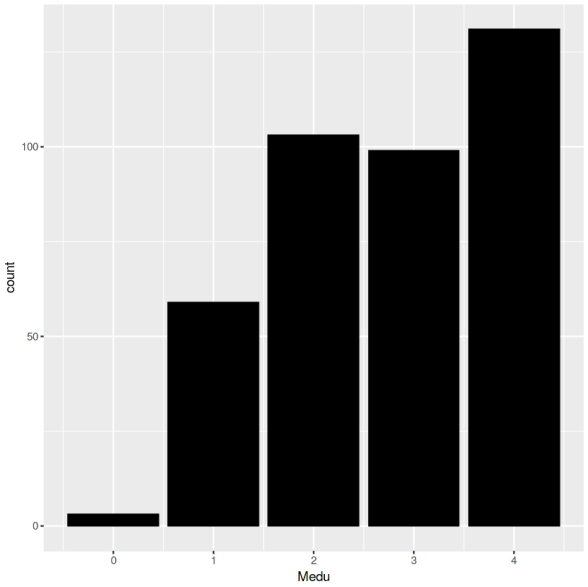

Figure. 11. Father education

ISSN 2723-6471

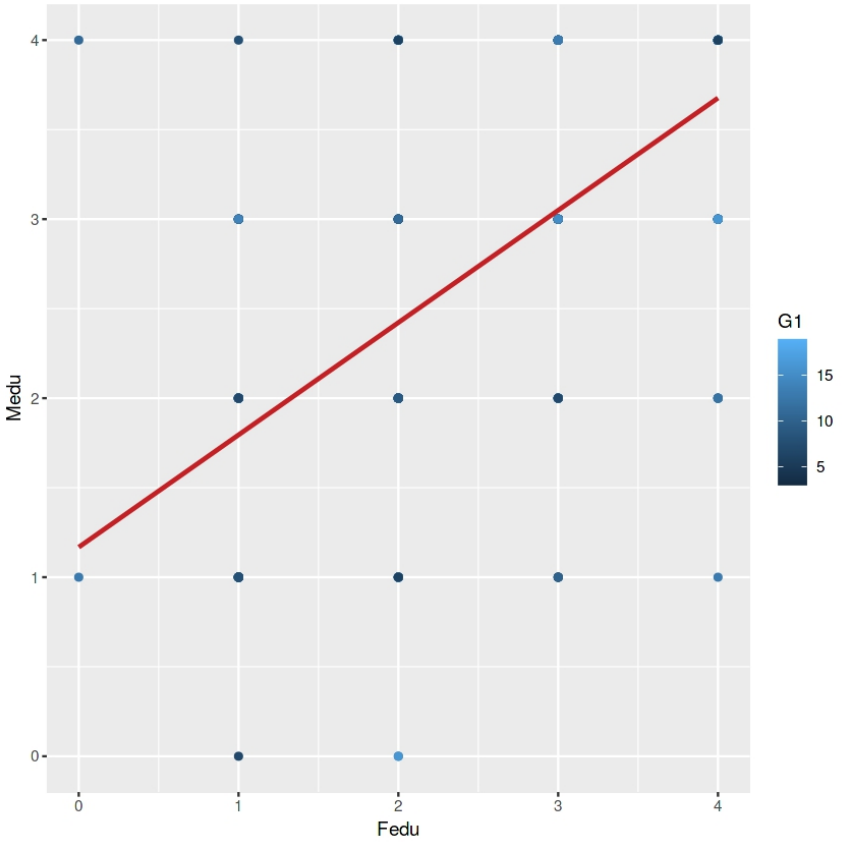

Figure. 12. G1 Father and Mother education

Like we expect, the highest the parent education is, the highest the kids score at school. Does this result mean something? Yes, students whose parents have a higher level of education may have an enhancement for learning, like more positive ability beliefs, a more solid work orientation, and maybe more effective learning strategies than children whose parents have lower levels of education.

ISSN 2723-6471

b) How family size contributes to students' academic performance?

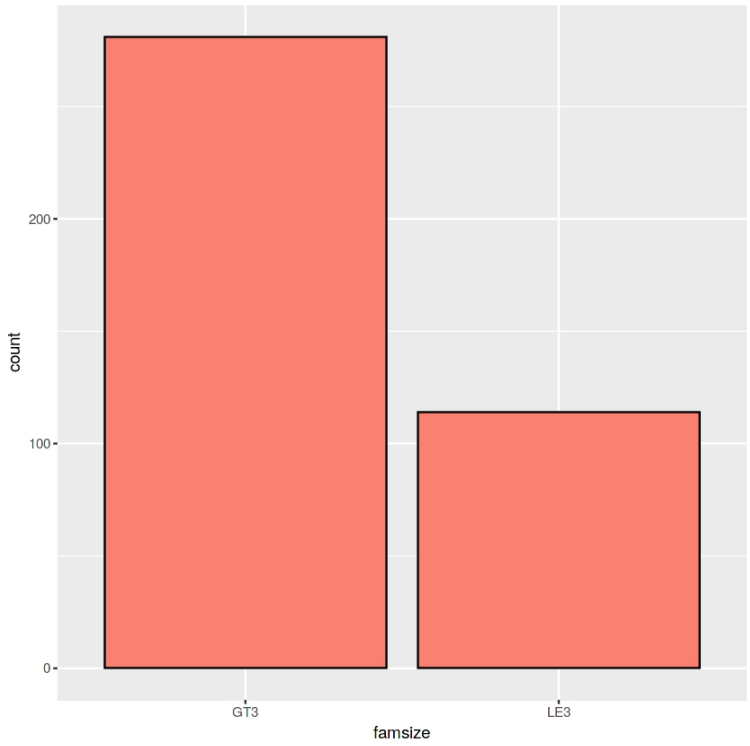

Figure. 13. Family size

From the graph above we identified that there are fewer students with family sizes less than 3 compared to families with family sizes greater than 3. According to some articles, The family size is the sideway or the other contribute to the success rate of the student in school, which mean that when the family is large, there is no adequate concentration on the child, it's like the parent need to contribute the exact amount of resources on every child and can't be focused only on 1 child.

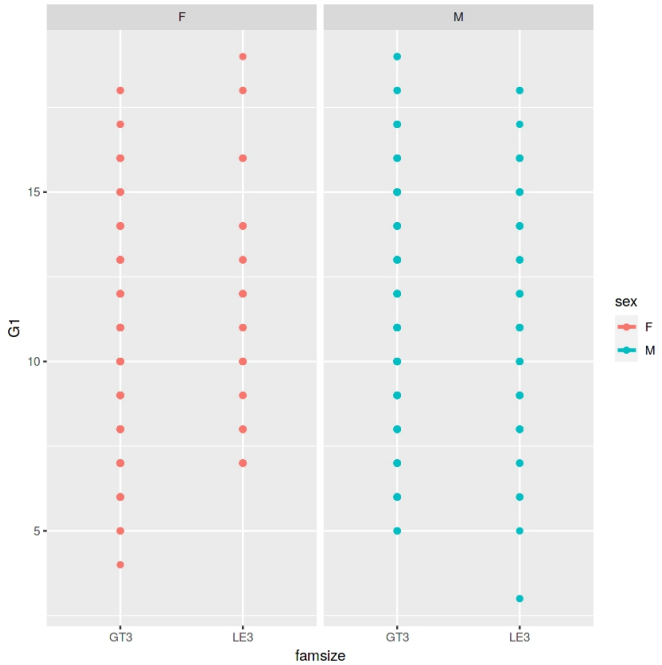

Figure. 14. G1 Family size

ISSN 2723-6471

c) Did kids of divorced parents perform lower in the class?

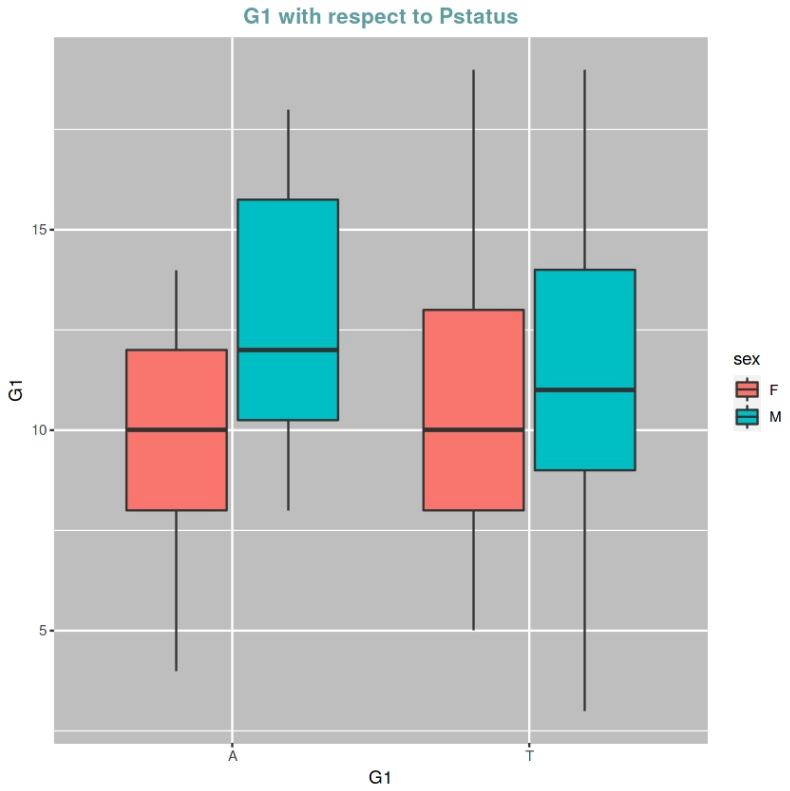

Figure. 15. G1 with respect to parents status

From the figures above, we can conclude that the median for female student G1 is at around 10 and for the male student it is around 12 and it's quite clear that students whose parents together score higher than students with divorced parents.

d) Did family relations affect student academic performance?

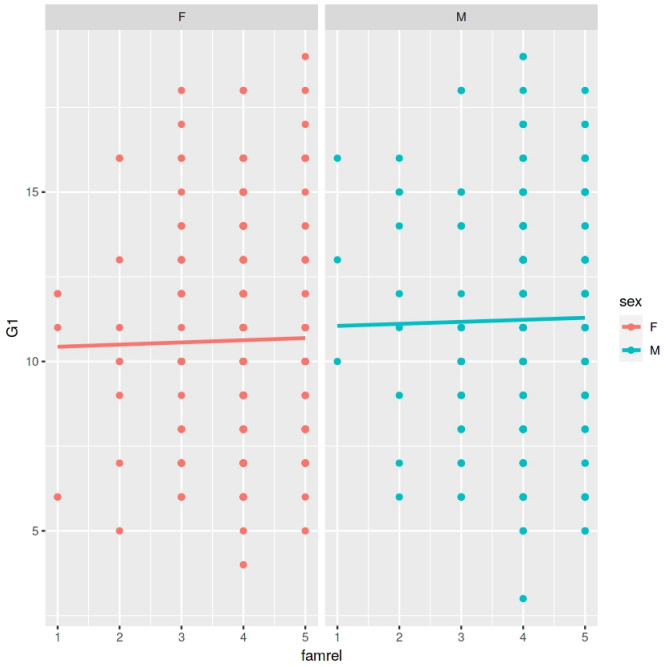

Figure. 16. Family relations

From the figure above we concluded that students with good family relations can perform better than the others, this happens because they have good communication and strong relationships and have better performance at school. and the graph above proved that student achievement increases with strong family relationships.

e) Does workday alcohol consumption affect the student's achievement?

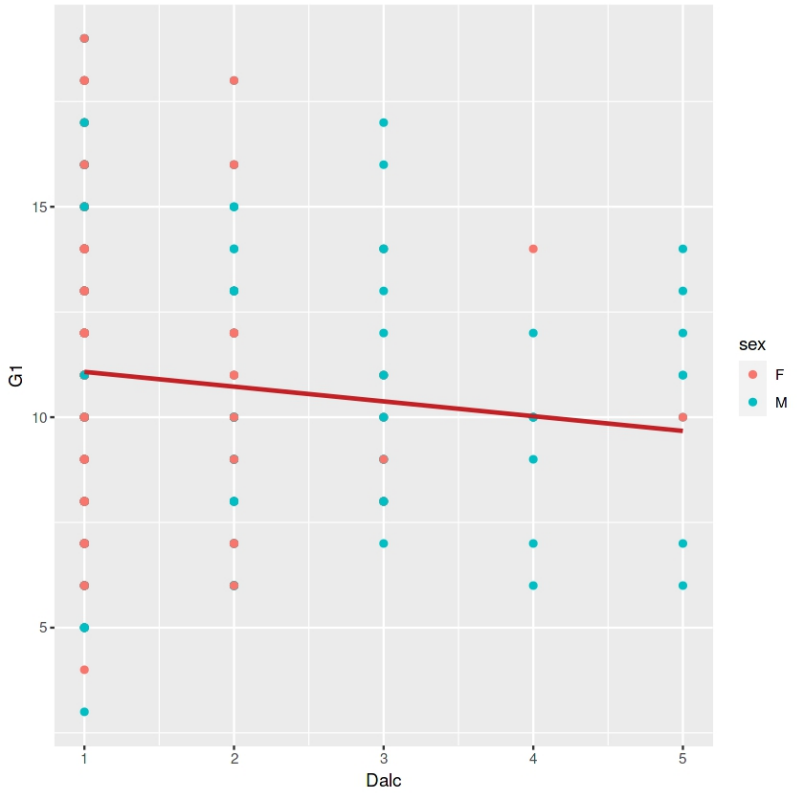

Figure. 17. Data Alcohol

Alcohol can indeed decrease student performance in school, An abuse of alcohol contributes to frequent confusion and memory loss. An inability to remember short terms and names can be caused by excessive alcohol consumption. And alcohol-abusing students often appear to lose their attention in class.

ISSN 2723-6471

4. Results and Discussion

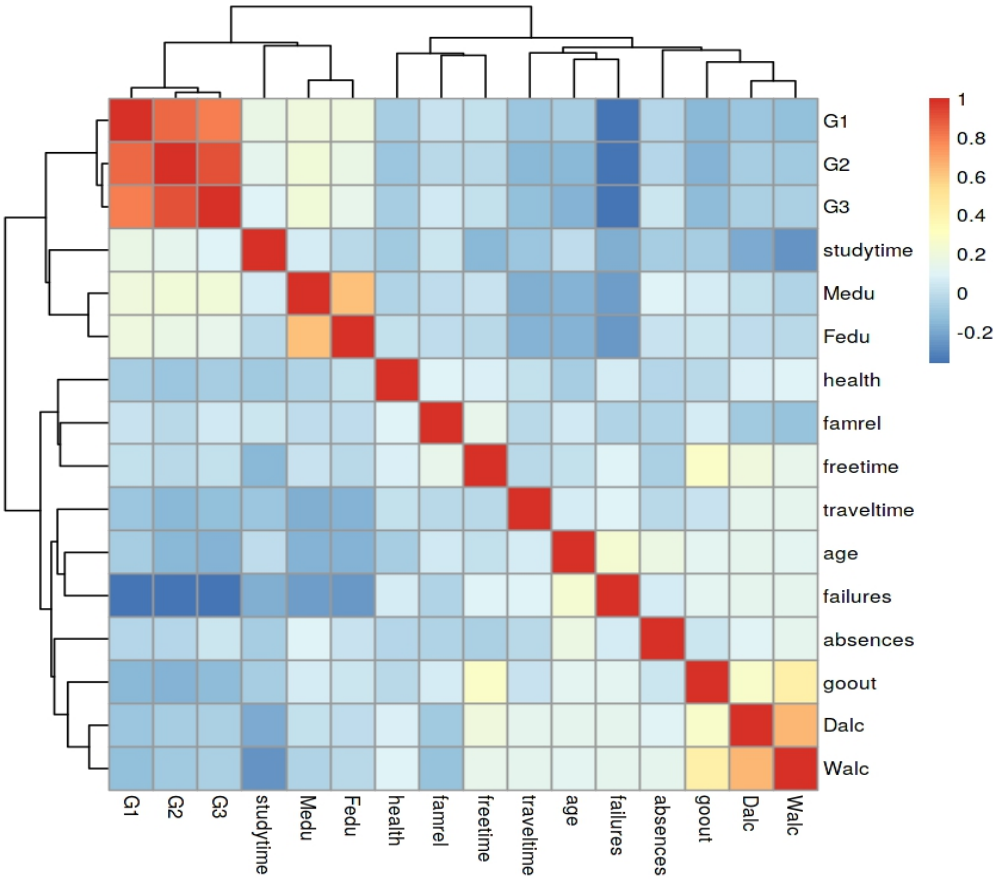

Figure. 18. Features relationship heatmap

From the heatmap above we can identify and analyse the relationship between the features we represent and how our variables connect into each other. It shows in the scale that our correlation varies between 1 (Positive correlation) and -1 (Negative correlation). Red color means the high correlated features, From the heatmap we can conduct that G1 and G3 can be classified having positive correlation on the other hand failures and study show negative correlation into it.

Finally after identifying correlation between every feature, we can start to predict the final grade (G3) using regression, It will go through 3 processes. First, we are going to Split or data partitioning. These steps are really important in order to divide data into Trained data and Test data. Second, we are going to Train the model, to create the good and stable prediction model we need to train and create our own formula to predict G3+ variables. Third, is model evaluation. In this process we are going to find out how our model is created and compare its prediction result.

On the first process data partitioning, we have completely Split the data into 2 data which is:

Table. 1. Data partition result

Trained Data	Test Data
279	116

Then we are going to build the model which is using linear regression model, in order to train the model on train data we create our own formula, to target and predict G3+ variables:

```
lin_mod=lm(G3~G1+G2+age+Fedu+Medu+Fjob+Mjob+famsize+sex+Pstatus+absences+famsize+Dalc+famrel
+traveltime, data=training)
```

Finally, the metric we are using to evaluate this model is MAE. What is MAE, Mean Absolute Error or known as MAE is some type of measure that measures the average magnitude of error in a set of prediction models [16], without considering their direction. It is the average of the absolute differences between prediction and real observation over the test sample, where all differences have equal weight [17-19]. from the MAE test we get 1.282333 as a value which is pretty low as we are expecting. Then how about checking if our model overfitted by computing the MAE training and comparing it to the MAE test. Train MAE we got a value 1.096498, it is almost the same value with the test value, which means our model didn't overfitting. After evaluating the model, we sure know the MAE value is low, so it means our model performs really well. Comparing our Predicted G3 and real G3 (table 2), it showed how close our prediction models were with a very small percentage of error.

Table. 2. Comparison between Predicted G3 and Real G3

	Predicted G3	**Real G3**
	<dbl>	**<int>**
1	5.139359	6
3	7.314031	10
4	14.027168	15
9	18.804189	19
10	15.379368	15
13	14.007247	14
14	10.461861	11
23	15.284462	16
24	13.083309	12
37	16.359049	18

5. Conclusion

In this study, we have proved that it is possible to create a prediction model for student grades. This means by all methods we also have been using the correct method and model to use, but there are still many more possibilities about another method in the future study. Whis kind research hopefully will help teachers, other scientists, etc to open their eyes that there is no limit to do this kind of research. The result of this study can be used to do some basic prediction using at least almost the same data set, it would be helpful as a teacher to predict their student grade. It will allow him to anticipate bad things and improvised something new to make their students achieve better grades.

References

[1] Romero, Cristóbal, and Sebastián Ventura. "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 (2010): 601-618.

[2] Baker, Ryan SJD, and Kalina Yacef. "The state of educational data mining in 2009: A review and future visions." *JEDM| Journal of Educational Data Mining* 1.1 (2009): 3-17.

[3] Dutt, Ashish, Maizatul Akmar Ismail, and Tutut Herawan. "A systematic review on educational data mining." *Ieee Access* 5 (2017): 15991-16005.

[4] Peña-Ayala, Alejandro. "Educational data mining: A survey and a data mining-based analysis of recent works." *Expert systems with applications* 41.4 (2014): 1432-1462.

[5] De Villiers, Ethel-Michele, et al. "Classification of papillomaviruses." *Virology* 324.1 (2004): 17-27.

[6] Ngai, Eric WT, Li Xiu, and Dorothy CK Chau. "Application of data mining techniques in customer relationship management: A literature review and classification." *Expert systems with applications* 36.2 (2009): 2592-2602.

[7] Ahmad, Fadhilah, Nur Hafieza Ismail, and Azwa Abdul Aziz. "The prediction of students' academic performance using classification data mining techniques." *Applied Mathematical Sciences* 9.129 (2015): 6415-6426.

[8] Bhardwaj, Brijesh Kumar, and Saurabh Pal. "Data Mining: A prediction for performance improvement using classification." *arXiv preprint arXiv:1201.3418* (2012).

[9] Antonie, Maria-Luiza, Osmar R. Zaiane, and Alexandru Coman. "Application of data mining techniques for medical image classification." *Proceedings of the Second International Conference on Multimedia Data Mining*. 2001.

[10] Thomas, Emily H., and Nora Galambos. "What satisfies students? Mining student-opinion data with regression and decision tree analysis." *Research in Higher Education* 45.3 (2004): 251-269.

[11] Ruß, Georg. "Data mining of agricultural yield data: A comparison of regression models." *Industrial Conference on Data Mining*. Springer, Berlin, Heidelberg, 2009.

[12] Bakar, Zuriana Abu, et al. "A comparative study for outlier detection techniques in data mining." *2006 IEEE conference on cybernetics and intelligent systems*. IEEE, 2006.

[13] Weber, Ben G., and Michael Mateas. "A data mining approach to strategy prediction." *2009 IEEE Symposium on Computational Intelligence and Games*. IEEE, 2009.

[14] Yeh, I-Cheng, and Che-hui Lien. "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients." *Expert Systems with Applications* 36.2 (2009): 2473-2480.

[15] Majumdar, Jharna, Sneha Naraseeyappa, and Shilpa Ankalaki. "Analysis of agriculture data using data mining techniques: application of big data." *Journal of Big data* 4.1 (2017): 20.

[16] Chai, Tianfeng, and Roland R. Draxler. "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature." *Geoscientific model development* 7.3 (2014): 1247-1250.

[17] Willmott, Cort J., and Kenji Matsuura. "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance." *Climate research* 30.1 (2005): 79-82.

[18] Franses, Philip Hans. "A note on the mean absolute scaled error." *International Journal of Forecasting* 32.1 (2016): 20-22.

[19] Akmal, "Predicting Dropout on E-learning Using Machine Learning," J. Appl. Data Sci., vol. 1, no. 1, pp. 29–34, 2020, [Online]. Available: http://bright-journal.org/Journal/index.php/JADS/article/view/6.