ADS-A Challenge: Your predictive Analysis

Dholon Akter

Dholon Akter

ADS-A

Team A: Ralf and Nick

11/27/2020

version-ii

Contents

Introduction

Although the educational level of the Portuguese population has improved in the last decades, the statistics keep Portugal at Europe's tail end due to its high student failure rates. In particular, lack of success in the core classes of Mathematics is extremely serious. On the other hand, the fields of Business Intelligence (BI)/Data Mining (DM), which aim at extracting high-level knowledge from raw data, offer interesting automated tools that can aid the education domain. The present work intends to approach student achievement in secondary education using BI/DM techniques. Recent real-world data (e.g. student grades, demographic, social and school related features) was collected by using school reports and questionnaires. This is a report of the individual challenge that was assigned to me during the ADS-A course. The project is about me making my personal predictive analysis. I had to apply Machine Learning using python within the Jupyter Notebook. In this report I will discuss what I put in my Business Proposal, what I did for my Exploratory Data Analysis, how I applied machine learning and what my conclusion and recommendations are.

Business Proposal

Most of the students are struggling to achieve better results in their final grade or want to succeed easily in every exam they face in their school. But the students or their parents do not know how to improve their grade and what is the reason behind failing the exam. For that reason, they keep struggling with their academic success. So here I have come up with the idea, if they know why they are not successful in their final exam and what is the reason for it, then they can be aware of those facts and move away from those activities. As a result, their performance will be improved gradually and get excellent academic results from their school.

Datasets

This data approaches student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school-related features) and it was collected by using school reports and questionnaires.

I chose a dataset because it contains academic and personal characteristics of the student as well as final grade. The task is to predict the final grade(G3) from the student information.

It is important to observe from the dataset, there are two columns which are the first two grades(G1 and G2) of students and it is difficult to predict the G3 without those data. But there are also other important data to predict the final grade.

Dataset link

https://www.kaggle.com/dipam7/student-grade-prediction?select=student-mat.csv

My prediction

I would like to predict the final grade based on different facts, for example using the internet, number of failures. Here G3 is the final grade and I can easily predict the final grade based on the calculation of two grades but there are other features which can be the reason to get high and low scores and pass or fail the exam. I will collect some features which I feel that the features can be the reason to predict the final grade or not

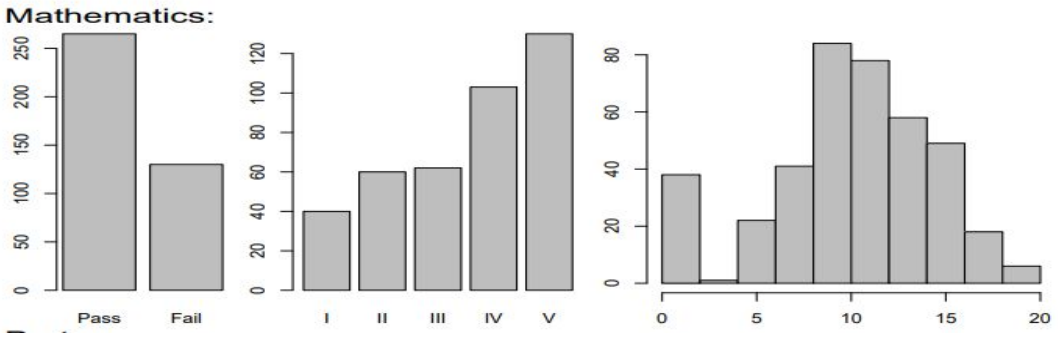
Exploratory Data Analysis

Here I want to explore the data how final grades can be affected by different factors.

So first I am going to look at the most relevant features of two periods of grade(**G1** and **G2**) to see the relationship of the final grade. so I made charts for both tableau and python to visualize the data. If necessary I have taken the charts from different sources.

Grade are classified of five levels in their exam

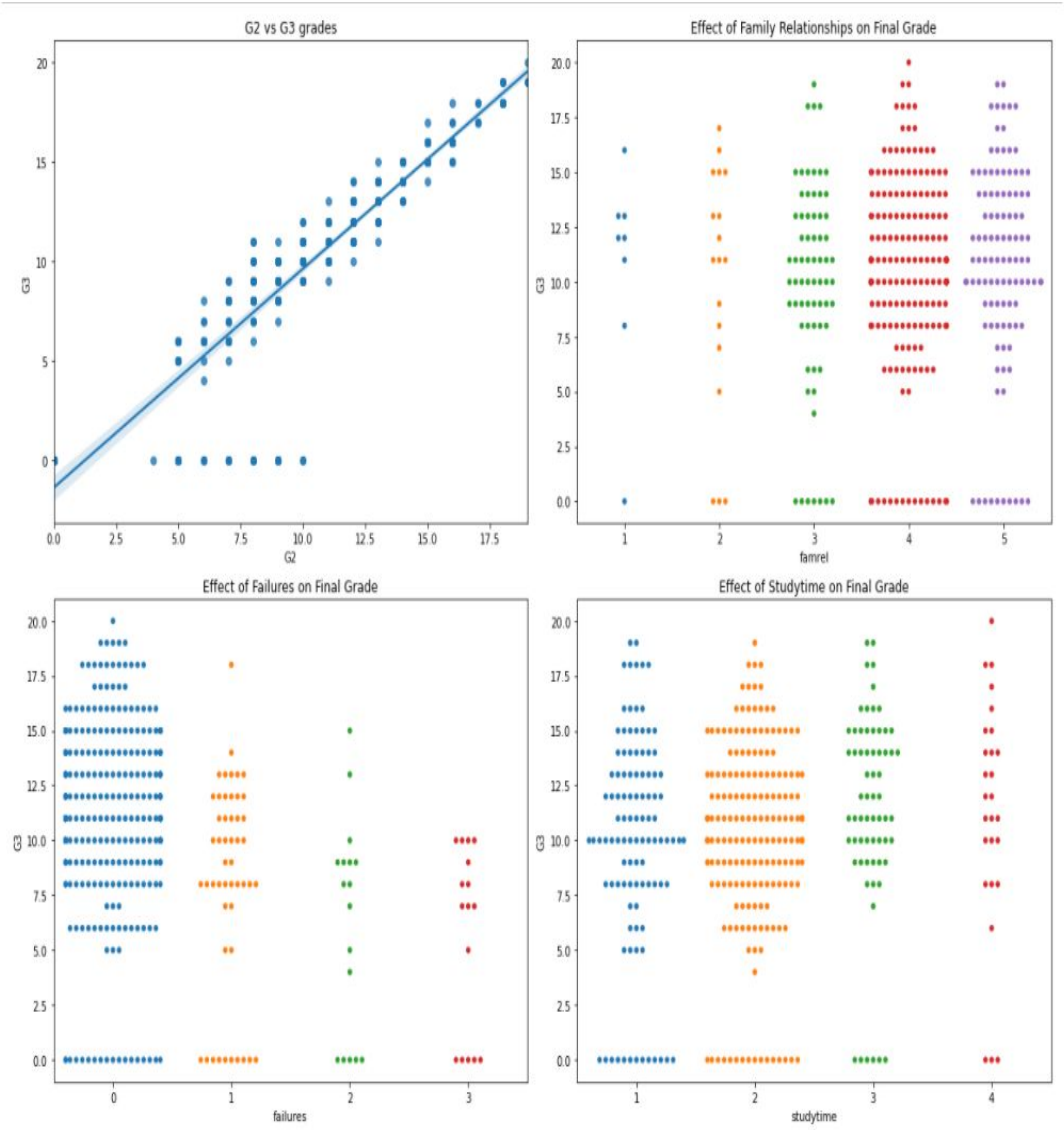
Country	I (excellent/very good)	II (good)	III (satisfactory)	IV (sufficient)	V (fail)
Portugal/France	16-20	14-15	12-13	10-11	0-9
Ireland	A	B	C	D	F



source : Dep. Information Systems/Algoritmi R&D Centre

From the source it is clear to say that , portugeese students of their grade classified with five level from very good or excellent to V-sufficient and with a numeric output the ranges between zero and twenty 100%. Above the calculation I can see how students are doing better in their exam.

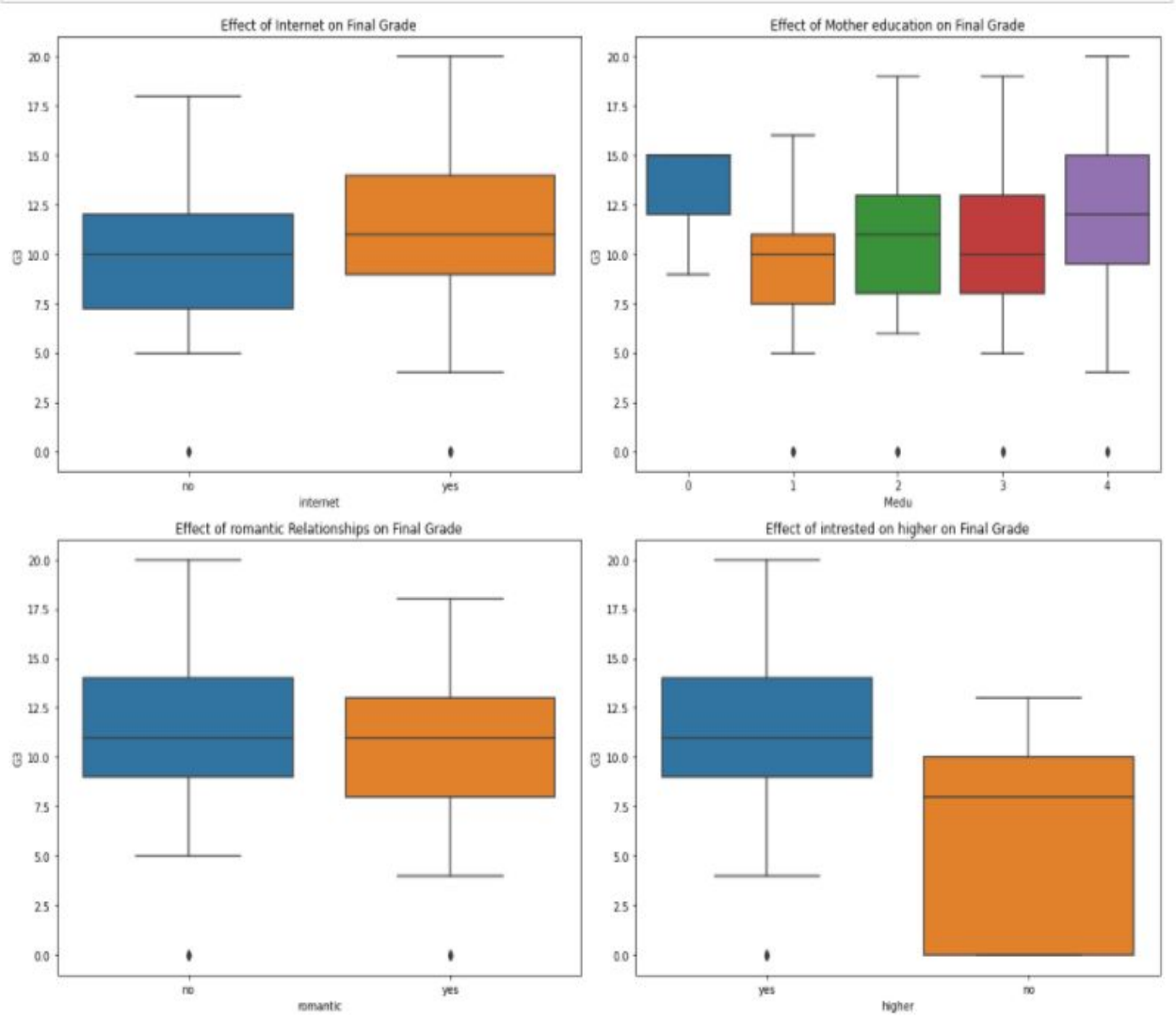
How Family relationship,failure,study time can affect the final grade



Dholon Akter-ADS A-Challenge

After looking at the data I found that these features are very relevant with final grade. These students who fail in class many times, the final grade of their exam is very low and they end up failing the exam in their final grade.

Students who like to do higher study, the chances of getting a higher result of their final exam.



Dholon Akter-ADS A-Challenge

It is clear from the chart that, the students who like to go higher study, the passing exam of their final grade is very high which is around more than 10 points, on the other hand the students do not have plan to go higher study, they end up fail in their final exam which is below 10 points.

Machine learning:

The machine learning algorithm I am going to use is logistic regression,linear regression, support vector machines and decision trees.. The target outcome will be the final grade.I will also, if possible, use KNN Regression. The project will succeed if I can get an accuracy score of at least 0.75.

How I applied machine learning and clean my data

In my demo section I showed Mr Ralf that my data is already clean and I do not have to clean the data. I used different algorithms for my prediction analysis where I got a 1.00 accuracy score in my linear regression which I am totally happy that there are no errors in my predictions.

Conclusion

My conclusion from the project is that I can , with a fairly accuracy, predict the final grade based on study time, family relationship, failures and higher study. If I were to continue this project, I would try to improve the linear regression model even more or try to combine this dataset with the dataset of the other countries so that the model will be able to predict the grade of all students all over the world.

My recommendation to Students, Parents or university is to contact me if they want me to predict the pass or fail rate of their students will probably get