Paula Castro & Gwak Seungmin
Group 10

1. Introduction

Education is a topic of major importance in today's societies. It is in our best interest that students achieve high performances, and many times this task is relegated to teachers. They must identify students with difficulties, and ensure their appropriate development. This is commonly done through heuristics or intuition. However, with the number of records available, data can help us make more informed decisions. We can use models to predict student performance, and effectively categorize them into different groups (low, middle, or high level) to more appropriately adapt to their needs, as is done in the Korean Middle School system. By developing a robust classifier system, and effectively assigning students to classes that better match their learning pace, we can provide a better learning experience to students, and maximize their academic performance. Also, we can look into the features these models are finding more relevant, and derive the factors that can aid in student success.

Also, we raise another question related to the recently introduced online learning environments. These allow students to remotely access class material, engage in discussion with peers or teachers, etc. All of this generates data that we previously didn't have. Can these new data sources help us create better predictions? If so, which attributes are the most relevant for success? Can these insights be translated to traditional learning systems for better classification?

Throughout the research, we find an array of interesting insights. For instance, we discover how including some attributes derived from the online learning platforms can be of great use. Also, we learn the importance of having well-rounded students that employ their time in different activities outside of school. This work-life balance seems to have a very positive repercussion.

2. Literature Review

In the literature, many papers trying to solve this same question can be found. The ways in which researchers have tackled this problem greatly varies. In [6] (Predicting Student Academic Performance using Support Vector Machine and Random Forest) binary classification model with SVM and RF algorithms was used to predict student performance. They use the similar dataset we use in the second analysis (We used the math grade dataset of Portuguese students, and they used the Portuguese grade dataset. They include the same feature.). Whereas they focus on the previous grades(G1, G2) and found that the previous grade has the most impact on the final grades, We're except the previous grades and focus to find other features which can make student performance better with building multi-class classification model.

In [5] (Mining Educational Data to Predict Student's academic Performance using Ensemble Methods), they made the student's performance predictive model with Artificial Neural Network (ANN), Naïve Bayesian (NB), and Decision tree (DT). In addition, they applied used Bagging, Boosting, and Random Forest (RF), which are the common ensemble methods. In this research, however, we focus on features that have strong impacts on prediction, so does more exploratory data analysis (EDA), and feature selection and engineering than theirs.

3. Problem Definition, Initial Exploration, and Experimental Evaluation

3.1 Online Vs Offline

With the recent online environment, several online features appeared in the educational field. Our first goal in this research was to identify the impact of the online learning environment on student performance prediction.

We found the dataset for finding which feature is useful for student performance prediction. The used dataset is publicly downloadable at Kaggle[1], and collected from LMS called Kalboard 360. The data include basic offline features and online participation/behavior features.

[1] https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data

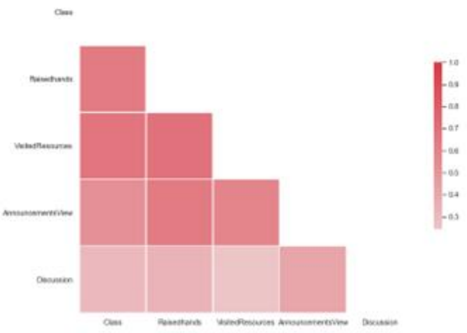
Our target is the total grade of a semester. Target classes are layered to low levels (0-69), medium levels (79-80), and high levels (90-100) depending on their grades, so we will use multiclass classification models.

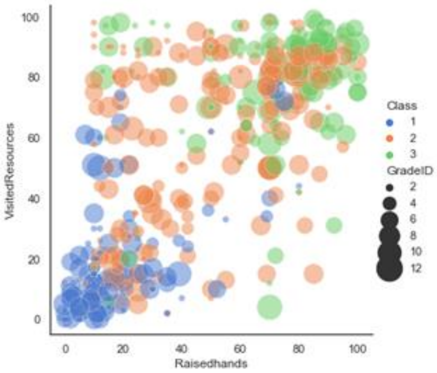
Features are divided into online and offline features. Online features include personal features, such as Gender and Nationality, educational features, such as School Levels, Topic, and Student Absence Days, and parent-related features, such as Parent responsible for the student. Offline features include discussion groups, visited resources, raised hands on class, and viewing announcement. The Kaggle download site contains all the attributes and their description. We're going to make a total of two models, one that uses only offline features and the other one that uses both online and offline features. By comparing the two we will find important features for student performance prediction.

3.1.1 EDA

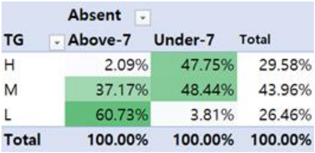
Before exploratory data analysis, we changed the data type of some features and target to numeric which weren't numeric but type changing makes EDA abundant. For example, We changed the target class from [Low, Middle, High] to [1, 2, 3]. We checked our dataset for common sense and got ideas for feature engineering.

We analyzed the online features first, which are our main concerns. Online features had a strong positive correlation with the target class except for discussion. The discussion had a relatively weak correlation. We made a relplot with Raised hands score and Visited Resources which have a higher correlation. Class distinction was more clear than the relplot of other online variable combinations.

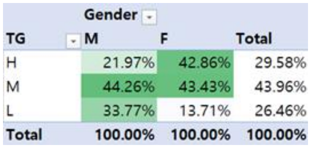




Regarding offline features, we checked the trends according to the bibliography. According to the bibliography, students with fewer absences have higher grades, and students have higher grades when their gender is female. Through the pivot table, we can see that those tendencies apply to our dataset as well. We predicted that being absent in particular could play a big role in distinguishing lower-level students. However, gender was not noticeable compared to other features.

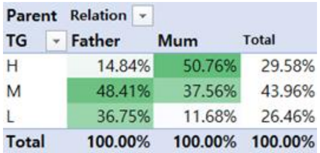


	Absent		
TG	Above-7	Under-7	Total
H	2.09%	47.75%	29.58%
M	37.17%	48.44%	43.96%
L	60.73%	3.81%	26.46%
Total	100.00%	100.00%	100.00%

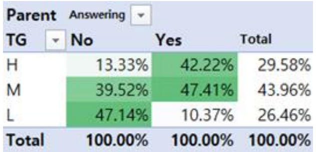


	Gender		
TG	M	F	Total
H	21.97%	42.86%	29.58%
M	44.26%	43.43%	43.96%
L	33.77%	13.71%	26.46%
Total	100.00%	100.00%	100.00%

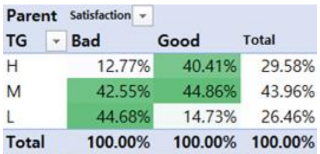
There were also some parent-related features in our dataset. The total grade distribution differed considerably depending on the value of the features. Student performance tends to be high when the parents who are responsible for the student are mothers, when the parents answer the survey, and when they are satisfied with the student's school. We were interested in feature engineering using these.



Parent	Relation		
TG	Father	Mum	Total
H	14.84%	50.76%	29.58%
M	48.41%	37.56%	43.96%
L	36.75%	11.68%	26.46%
Total	100.00%	100.00%	100.00%



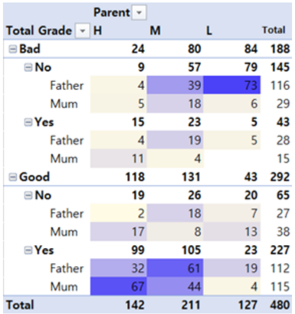
Parent	Answering		
TG	No	Yes	Total
H	13.33%	42.22%	29.58%
M	39.52%	47.41%	43.96%
L	47.14%	10.37%	26.46%
Total	100.00%	100.00%	100.00%



Parent	Satisfaction		
TG	Bad	Good	Total
H	12.77%	40.41%	29.58%
M	42.55%	44.86%	43.96%
L	44.68%	14.73%	26.46%
Total	100.00%	100.00%	100.00%

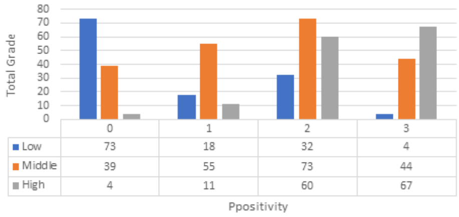
3.1.2 Feature Engineering

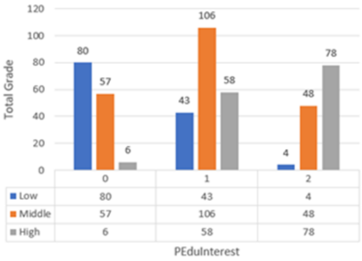
From EDA, we get some more understanding of our data and it gives us an idea about feature engineering.

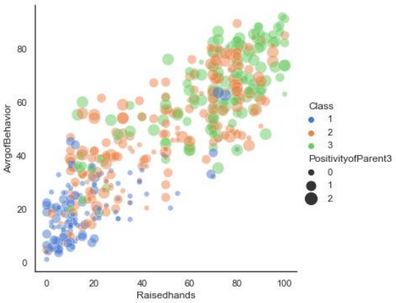


	Parent			
Total Grade	H	M	L	Total
⊟ Bad	24	80	84	188
⊟ No	9	57	79	145
Father	4	39	73	116
Mum	5	18	6	29
⊟ Yes	15	23	5	43
Father	4	19	5	28
Mum	11	4		15
⊟ Good	118	131	43	292
⊟ No	19	26	20	65
Father	2	18	7	27
Mum	17	8	13	38
⊟ Yes	99	105	23	227
Father	32	61	19	112
Mum	67	44	4	115
Total	142	211	127	480

1) **Parent's Positivity and Parent's Educational Interest –** The parent-related features are binomial categorical or boolean type. We translate Mum and Yes, Good to 1 and the other is 0. We first created PPositivity by adding all. However, PPostivity can't make a story of it so we modified the method. We created PEduInterest with the value of adding Relation and Answering. It has similar distribution with Satisfaction, so we replace Satisfaction with PEduInterest.







2) **Average of Behavior –** We created and calculated the average of feature participation scores. It shows the overall online participation. With the engineered feature, it is better to compare with other online features' tendency and predict the performance.

The size of the spot of instance in the left graph is Parent's Educational Interest we introduced before, and the y axis of the graph is Average of Behavior.

3.1.3 Methodology

We built an online model and an offline model, respectively, and in this process, we focused on feature selection and combining the features.

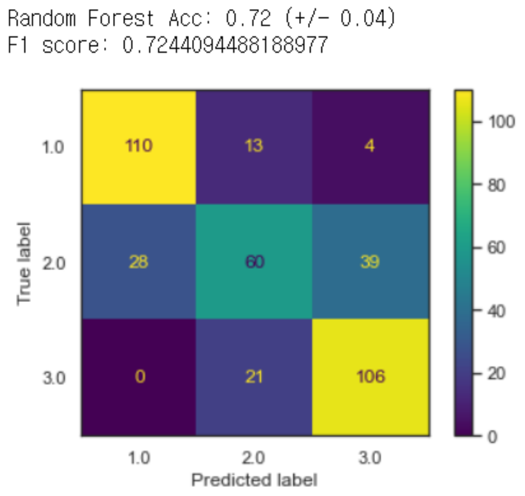
Considering that it is a multi-class classification, we used F1 score, not AUC, and drew a confusion matrix to better understand the model's performance. With some running trials, we found that rebalancing the dataset increases model performance, so we tested several rebalancing techniques (Undersampling, Oversampling, SMOTE, and Tomek Links) one by one. We also used several feature selection techniques (Stepwise Recursive Backwards Feature removal, Wrapper Select via model, and Univariate Feature Selection – Chi-squared) to identify the selected features from them. We referred to these to find the best feature combination (or a subset of features).

3.1.4 Results

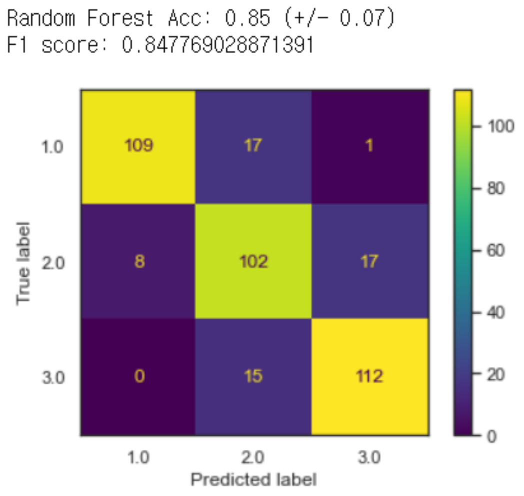
When the experiment setup was ready, each condition was executed and the accuracy, F1 score, and runtime were recorded. In both online and offline models, the random forest algorithm was the most suitable, and the second-best algorithm of the online model was the gradient boosting and one of the offline model was the support vector machine. When Undersampling was applied to both models, the performance of the model improved. This seems because the application of Undersampling reduces ambiguous instances which are difficult for the models to classify high or not in this dataset.

When using feature selection methods, nationality was included in the results of two of the three methods. However, only Jordan and Kuwait have nationality with more than 30 instances in the dataset. Also, some nationalities have extreme distributions of the target. Although performance is improved, it is not a statistically significant improvement, nor is it helpful for generalization, so we excluded the feature.

As a result, we were able to improve accuracy by 13%p and F1 score by 0.12 points when using online features. There were particularly significant improvements in distinguishing between high and medium levels. The left confusion matrix is the offline model, and the right one is the online model.



```
Random Forest Acc: 0.72 (+/- 0.04)
F1 score: 0.7244094488188977
```



```
Random Forest Acc: 0.85 (+/- 0.07)
F1 score: 0.847769028871391
```

3.1.5 Conclusions

We were able to extract the following insights from the above results.

First, online behavior variables improve performance for classifying high and non-high levels. Among the online features, features with a high correlation coefficient with the target and a high importance score are visited resources. Basically, online features were things whose character could be explained in terms of traditional education. Visited resources can be translated as "review," and we can see the importance of reviewing. The next important variable is raised hands. However, it is a little too much for us to clearly identify the causal relationship between the number of presentations and grades.

Features about parents also provided important information for predicting student performance. This means that parents are important to student performance. However, this is not a causal relationship, so it does not mean that a student's mother should be more responsible for the student, respond more faithfully to the survey, or be satisfied with the school to improve student performance.

Through the first analysis, we found that features related to review, presentation, and attendance are important for predicting student performance. However, the presentation was difficult to determine the causal relationship, and there were features that students could not control, such as features related to parents. Therefore, it remains a question for us what features students should care about and can use to improve their grades.

3.2 Diving Deeper

Having analyzed how online and offline measures of student engagement can affect our predictive power on student performance, we now aim to dive deeper into the problem. We decided to look for a new dataset, in which we could analyze more aspects of students' behavior and environment, and test the effect these had on our model performance. Ideally, we aimed to find the attributes that distinguished high-performing students, and see if these could be adopted by other students to boost their academic outcomes. Also, by being able to build a robust model, we could provide an objective way to classify students into different classes according to their level (as is done in Korean high schools). This creates for homogeneous classes, where students' necessities are taken into account, therefore possibly maximizing learning efficiency. All in all, our research project would help students understand what factors impact their performance, and schools in creating tailored programs for students to incentivize a better learning experience.

The dataset used is publicly downloadable at Kaggle[2], and includes information about student performance in 2 Portuguese high schools, both in Portuguese and math subjects. In this analysis, we will only focus on the latter. The features were collected through reports and questionnaires, and cover a wide array of aspects, ranging from social to school-related features. The Kaggle download site contains all the attributes and their descriptions.

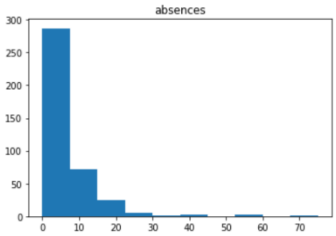
As targets, we have also been provided the attributes 'G1', 'G2', and 'G3', which correspond to the first and second-period grades, and the final grade respectively. For this analysis, we will only be using 'G3'. This is because we are interested in predicting student performance at the beginning of the school year. Also, all of these values are highly correlated, so we can assume that student performance will be stable throughout the year. For both of the analyses in this research paper to be consistent, we will be transforming the numeric 'G3' attribute into a factor of 3 levels, 'L', 'M', and 'H'. These represent low, medium, and high achieving students respectively.

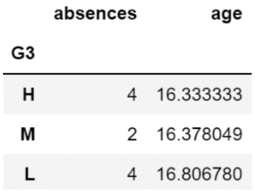
3.2.1 EDA

Before diving into model building, we needed to understand our data and identify possible problems. We performed a series of diagnostic plots, to check the distribution of our data and understand the information each variable was conveying. For numeric variables, we made histograms, correlation plots, and pivot tables. For categorical data, we created bar charts and used pivot tables to see how these groups affected our target variable. All of this started uncovering interesting nuances in the dataset and gave us some ideas for future data engineering. Some of the **initial questions** we wanted to answer were the following: Does more '*studytime*' relate to higher achievements? Do students receiving support of any type perform better? Does alcohol consumption affect performance? Do students that go out more often have lower grades?
Firstly, we analyzed the numeric variables, age, and absences. To grasp whether these 2 factors influenced student performance, we created a pivot table. Each row represented the different classes of students according to G3. The columns are the median number of absences and the average age of

[2] https://www.kaggle.com/datasets/barkhaverma/student-performance

students in each group. We used the median for the absences column because there are many outliers in this feature, which would highly skew the group means.





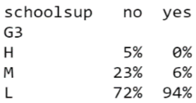
G3	absences	age
H	4	16.333333
M	2	16.378049
L	4	16.806780

From the pivot table, we can see that age may be a factor that distinguishes students, where lower-achieving students tend to be older. However, we don´t know if this difference is significant, so we continue our analysis. Regarding the number of absences, no pattern seems to arise.

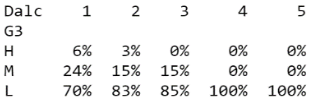
Next, we set out to explore the categorical variables. In the first instance, we noticed that many of the variables were very unevenly distributed. We created some pivot tables, to compare the groups across different performance levels. Due to the imbalances, we normalized the value per group, to better interpret the results.

In this step, we realized that some attributes added no value to our model. The distribution of our target class was essentially the same in all the groups, so they had no discriminative power. We opted to remove these variables or reshape them into more informative features.

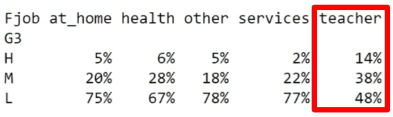
However, other variables seemed to have some very discriminative values. These helped us formulate some interesting questions, and gave us some ideas as to how to reshape our features to be more informative. An example of this is shown below. If we look at the Fjob variable, we can see that the different categories don´t seem to be very different concerning the distribution between high, medium, and low students within them. However, this changes when Fjob = teacher. These students have higher prior probabilities of belonging to the higher achieving groupings. Therefore, our model may benefit from a single binary variable Fteacher (0 if the parent is not a teacher, and 1 otherwise).



```
schoolsup   no  yes
G3
H           5%   0%
M          23%   6%
L          72%  94%
```



```
Dalc     1    2    3     4     5
G3
H       6%   3%   0%    0%    0%
M      24%  15%  15%    0%    0%
L      70%  83%  85%  100%  100%
```



```
Fjob at_home health other services teacher
G3
H         5%     6%    5%       2%     14%
M        20%    28%   18%      22%     38%
L        75%    67%   78%      77%     48%
```

3.2.2 Feature Engineering

After EDA, we created a better understanding of our data and the relationship between variables and targets. We had formed some intuition about possible feature transformations that could be more informative.

1) Alcohol – The dataset contained 2 variables related to alcohol consumption, weekday and weekend averages. These were highly correlated, and therefore we could create a new feature, Average weekly alcohol consumption.

2) Support – In the previous data exploration, we saw that school support was the only support type that seemed to impact academic outcomes. Therefore, we create a single binary feature, *Support*, that indicated whether the student received school support (1) or otherwise (0). We can see below the new discriminative power of this variable.

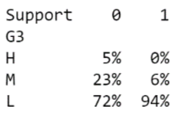
3) Social Life – It seemed as if '*goout*' variable could be relevant if treated correctly. We hypothesized that it would be more informative if we measure the 'social life' by taking the ratio between '*freetime*' to ‘*goout*’. This way we could see how much the student was investing in going out relative to the amount of spare time he/she had. At first glance, it looked like the new variable didn´t bring great improvements.

4) Study – Similar to above, we hypothesized that study time would also be more informative when we took the ratio of it with respect to ‘*freetime*’. By doing this, we would be able to measure how much of the spare time students spent studying. This new feature looked very promising. There was a big distinction between the highest achieving group and the other 2.

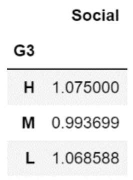
5) Repeaters – Throughout the analysis, we noticed that age and failures seemed to be related to lower-achieving students. We decided to merge these 2 variables into a new one, Repeaters. Those students over 16 years old and with 1 or more failures were classified as so. They seemed to be more prone to belong to the ‘L’ group.

6) Guardian – Similar to what happened with the types of support, we saw that the guardian variable only had a significant impact on ‘*G3*’ when its value was ‘*other*’. Therefore, we regrouped the values in this group into ‘*parent*’ or ‘*other*’. Students where ‘$guardian = other$’ seemed more prone to belong to the ‘L’ group.

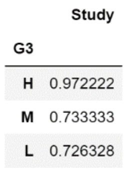
7) Fteacher – As explained above, the ‘Fjob’ variable showed significant group differences between it being ‘*teacher*’ versus any other value. Therefore, we created the new binary variable ‘*Fteacher*’, indicating whether the father is employed as a teacher (1) or not (0). A positive response to this seemed to be related to better performance.



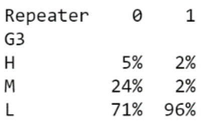
```
Support    0    1
G3
H         5%   0%
M        23%   6%
L        72%  94%
```



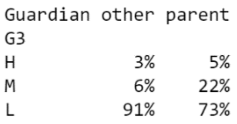
	Social
G3	
H	1.075000
M	0.993699
L	1.068588



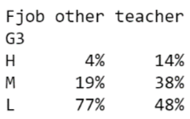
	Study
G3	
H	0.972222
M	0.733333
L	0.726328



```
Repeater     0    1
G3
H           5%   2%
M          24%   2%
L          71%  96%
```



```
Guardian other parent
G3
H            3%     5%
M            6%    22%
L           91%    73%
```



```
Fjob other teacher
G3
H       4%     14%
M      19%     38%
L      77%     48%
```

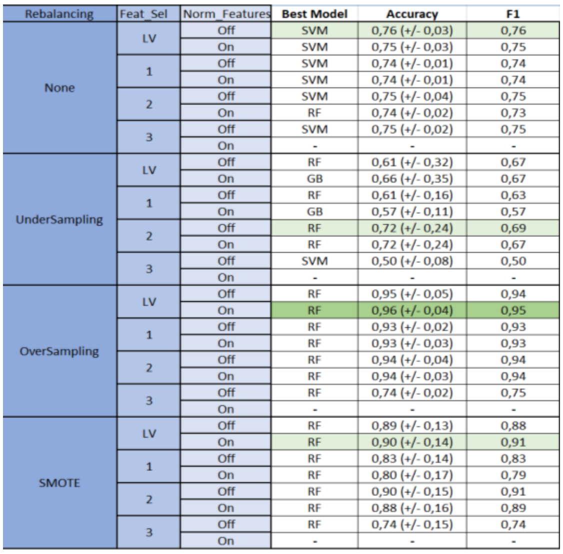
3.2.3 Methodology

After feature engineering, we had our dataset ready to be modeled. To do this, we took into account that we were dealing with a multiclass classification problem. This impacted the algorithms we chose to compare (Decision Trees, Random Forest, Gradient Boosting, MLP, and SVMs), and the metrics we used to evaluate performance (Accuracy and F1 score). We also decided to plot the confusion matrices to see in which part our algorithms were failing most and act accordingly.

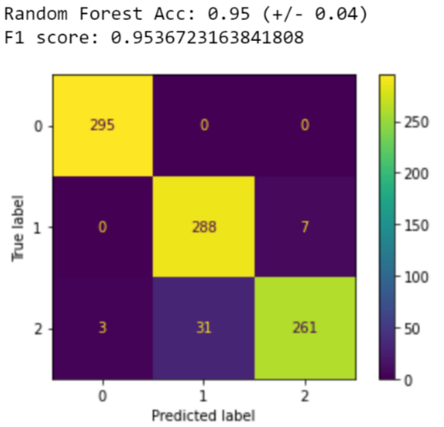
When running the first set of models, we realized that class imbalance in our target was a big problem, so we included some rebalancing techniques in the pre-processing section. These were under-sampling, over-sampling, and SMOTE. We also tested out how feature normalization affected our results, as well as different types of feature selection (Low variance filter, Stepwise Backwards Removal, Wrapper Selection, or Univariate Selection).

3.2.4 Results

Once we had the experimental setting ready, we ran each of the conditions and recorded the best achieving model, together with its accuracy and F1 score. This helped us locate the best-achieving model, a Random Forest. Oversampling was applied to the data, as well as using the Low Variance filter, and feature normalization. Model performance is displayed below. We must point out this analysis was conducted for the original data too, and the results were only slightly worse. However, this model involved more predictors and was more difficult to interpret (all features had low feature importance scores), so we kept the one with feature engineering.

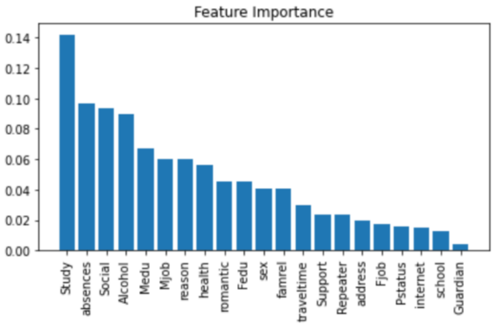


Rebalancing	Feat_Sel	Norm_Features	Best Model	Accuracy	F1
None	LV	Off	SVM	0,76 (+/- 0,03)	0,76
		On	SVM	0,75 (+/- 0,03)	0,75
	1	Off	SVM	0,74 (+/- 0,01)	0,74
		On	SVM	0,74 (+/- 0,01)	0,74
	2	Off	SVM	0,75 (+/- 0,04)	0,75
		On	RF	0,74 (+/- 0,02)	0,73
	3	Off	SVM	0,75 (+/- 0,02)	0,75
		On	-	-	-
UnderSampling	LV	Off	RF	0,61 (+/- 0,32)	0,67
		On	GB	0,66 (+/- 0,35)	0,67
	1	Off	RF	0,61 (+/- 0,16)	0,63
		On	GB	0,57 (+/- 0,11)	0,57
	2	Off	RF	0,72 (+/- 0,24)	0,69
		On	RF	0,72 (+/- 0,24)	0,67
	3	Off	SVM	0,50 (+/- 0,08)	0,50
		On	-	-	-
OverSampling	LV	Off	RF	0,95 (+/- 0,05)	0,94
		On	RF	0,96 (+/- 0,04)	0,95
	1	Off	RF	0,93 (+/- 0,02)	0,93
		On	RF	0,93 (+/- 0,03)	0,93
	2	Off	RF	0,94 (+/- 0,04)	0,94
		On	RF	0,94 (+/- 0,03)	0,94
	3	Off	RF	0,74 (+/- 0,02)	0,75
		On	-	-	-
SMOTE	LV	Off	RF	0,89 (+/- 0,13)	0,88
		On	RF	0,90 (+/- 0,14)	0,91
	1	Off	RF	0,83 (+/- 0,14)	0,83
		On	RF	0,80 (+/- 0,17)	0,79
	2	Off	RF	0,90 (+/- 0,15)	0,91
		On	RF	0,88 (+/- 0,16)	0,89
	3	Off	RF	0,74 (+/- 0,15)	0,74
		On	-	-	-



```
Random Forest Acc: 0.95 (+/- 0.04)
F1 score: 0.9536723163841808
```

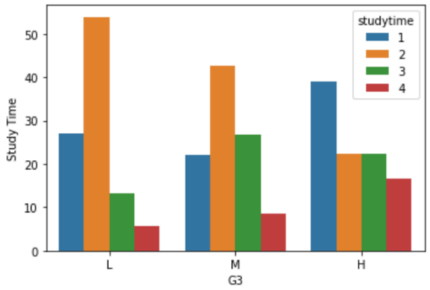
One of the advantages of using Random Forests is the fact that we can derive the feature importance's, to better understand how our model is taking decisions. For the model above, the results were very interesting, as it turned out many of the important features for classification were those that we had engineered (Study, Social, Alcohol…).



3.2.5 Conclusions

The results above give us a very interesting insight into which variables seem to be relevant to determining student success. For instance, we have collected evidence supporting some of the common heuristics (alcohol consumption and absences being detrimental to performance), while also challenging others (receiving support will improve performance). Now we are finally in a place to answer the questions we posed during EDA.

- *Does more 'studytime' relate to higher achievements?*

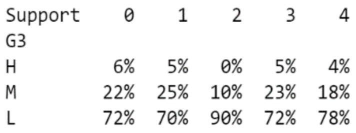


Although the proportion of students dedicating more time to studying is higher in the 'H' group than in the others, we can see how the majority of higher-achieving students don´t employ much time studying. However, when considering the ratio between time spent studying and free time, it becomes very apparent that this feature *is* discriminative between groups.

Therefore, we can conclude that study time itself is not informative. Rather, we should look at the proportion of free time spent studying. This is higher in students from group 'H'. We can therefore infer that even though from the above graph it looks like high achieving students spend less time studying, the reality is that they just have less time, but spend more of it studying. For example, we see the social life ratio is the highest in the 'H' group too. We can hypothesize this is making them more 'well-rounded' students and boosting their performance.

- *Do students receiving support of any type perform better?*

To answer this question, we created a new categorical variable, taking values from 0 to 4. It represents the type of support given to a student (0- No support, 1- family support, 2-school support, 3- paid support, 4- combined support). We can look at the distribution:



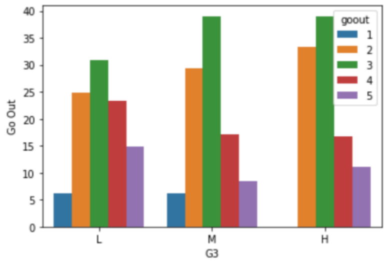
```
Support    0    1    2    3    4
G3
H         6%   5%   0%   5%   4%
M        22%  25%  10%  23%  18%
L        72%  70%  90%  72%  78%
```

As explained before, only school support seemed to impact the student distribution across 'G3' groups. This explains our feature engineering choices above. We can conclude that receiving school support is associated with lower-achieving students. Regarding the rest of the support types, not enough evidence is available.

- *Does alcohol consumption affect performance?*

As seen in the visualizations above, it seems that those students with higher weekly alcohol consumption tend to perform worse.

- *Do students that go out more often have lower grades?*



It looks like the distribution between groups is similar. However, 'H' students seem to go our more, on average. This ties in with our conclusions regarding studytime. It suggests that higher-achieving students have a better balance between social, scholarly, and other activities.

4. Future Work

There are some improvements that could have been made to the experimental design and execution of this research. Firstly, we acknowledge that the amount of data was limited, maybe biasing our results in certain directions. It is for this reason that in the future we would like to create a more comprehensive dataset, in which schools from different parts of the world, and different students from varying socio-demographical backgrounds are considered. Also, regarding the second part of the analysis (2 Portuguese high schools), the data collection process was through surveys passed down to the students, where many of the questions were categorical (scales between 1 to 5) instead of real measurements. We consider that having a more standardized and objective way of collecting the data would benefit the research conclusions. For example, measuring the total amount of time spent studying, rather than grading it on a scale of 1 to 5. Lastly, we also consider more data could have been collected from the online platforms, such as the amount of time spent on the LMS, how much time before the assignment due dates papers are submitted, etc. This can be an interesting future field of study.

5.Bibliography

[1] Cortez, P., & Silva, A. (2022). *Using data mining to predict secondary school student performance*. Retrieved May 18, 2022, from http://www3.dsi.uminho.pt/pcortez/student.pdf

[2] Martins, M.V., Tolledo, D., Machado, J., Baptista, L.M.T., Realinho, V. (2021). Early Prediction of student's Performance in Higher Education: A Case Study. In: Rocha, Á., Adeli, H., Dzemyda, G., Moreira, F., Ramalho Correia, A.M. (eds) Trends and Applications in Information Systems and Technologies. WorldCIST 2021. Advances in Intelligent Systems and Computing, vol 1365. Springer, Cham. https://doi.org/10.1007/978-3-030-72657-7_16

[3] Javier López-Zambrano , Juan Alfonso Lara Torralbo , and Cristóbal Romero. (2021). *Early Prediction of Student Learning Performance Through Data Mining: A Systematic Review*. Psicothema. https://www.psicothema.com/pdf/4692.pdf

[4] Gamazo A and Martínez-Abad F (2020) An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques. Front. Psychol. 11:575167. doi: 10.3389/fpsyg.2020.575167

[5] Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. International Journal of Database Theory and Application, 9(8), 119-136. DOI:10.14257/IJDTA.2016.9.8.13

[6] Leena H. Alamri, Ranim S. Almuslim, Mona S. Alotibi, Dana K. Alkadi, Irfan Ullah Khan, and Nida Aslam. (2020). *Predicting Student Academic Performance using Support Vector Machine and Random Forest*. In 2020 3rd International Conference on Education Technology Management (ICETM 2020), December 17–19, 2020, London, United Kingdom. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3446590.3446607