Predicting Student's Academic Performance

Nhut-Truong Le Nguyen and Quoc-Thang Vo and Cam-Nhung Bui.

I. DATASET REVIEW

A. Dataset information

This data approach 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. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language (por).

B. Dataset description

II. APPLY DATASET FEATURES ON SUTDENT'S PERFORMANCE

Using the final exam performance, classifying these students into three categories, "good", "fair", and "poor. Then, pointing out a few features that have significant influence on students' final performance, including:

- Romantic Status.
- Alcohol Consumption.
- Parents Education Level
- Frequency Of Going Out
- Desire Of Higher Education
- Living Area.

Finally, leveraging available features, Using various machine learning models to predict students' final performance classification and have compared models performance based on one-out sample accuracy score.

A. Dataset description

Some of attributes included in the dataset.

- address - student's home address type (binary: 'U' - urban or 'R' - rural)
- Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
- Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
- studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
- romantic - with a romantic relationship (binary: yes or no)
- goout - going out with friends (numeric: from 1 - very low to 5 - very high)
- Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
- Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
- absences - number of school absences (numeric: from 0 to 93)

  # These grades are related with the course subject, Math or Portuguese:
- G1 - first period grade (numeric: from 0 to 20)
- G2 - second period grade (numeric: from 0 to 20)
- G3 - final grade (numeric: from 0 to 20, output target)

See UCI Data Set description[1] for more infomation

B. Final grade distribution

The final grade plays a major role in comparing the efficient between different models. The closer to the final grade, the more efficient the model is.Students who had final grades between 15 and 20 will be classified as "good". For those whose final grades from 10 to 14 will be considered as "fair".And the rest will be classified as "poor". Related attributes: final_grade.

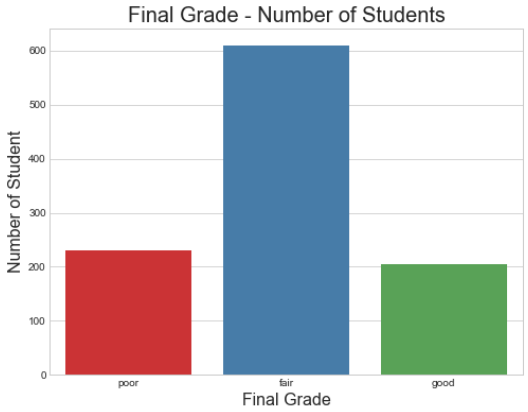

Fig. 1. Final grade distribution

C. Final grade by romantic status

Comparison academic performance between students in a romantic relationship and not in a relationship. Compare between final_grade and romatic attributes in dataset.

D. Final grade by alcohol consumption

The link between alcohol consumption and students performance. Data retrieved from final_grade and weekly_alcohol_usage in dataset.

E. Final grade by parents education level

The influence of parents education level on students performance related attributes: mother_education, father_education Result: Parents' education level has a positive correlation with students' final score. Comparatively, mother's education level has bigger influence than father's education level.

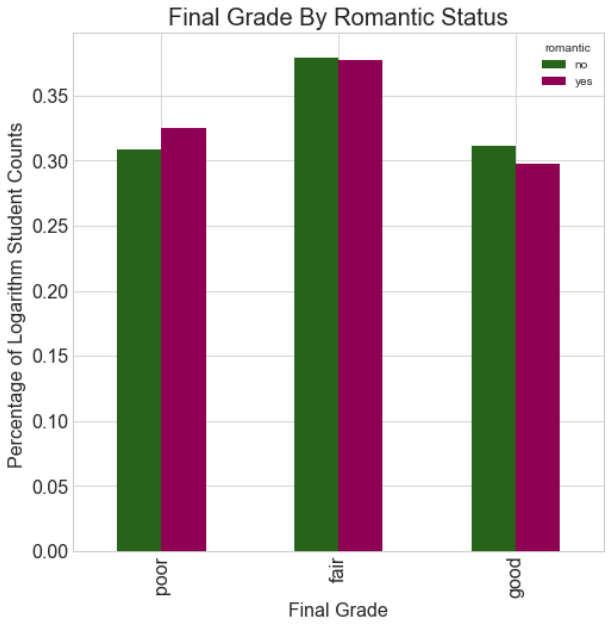

Fig. 2. Final grade by romantic status

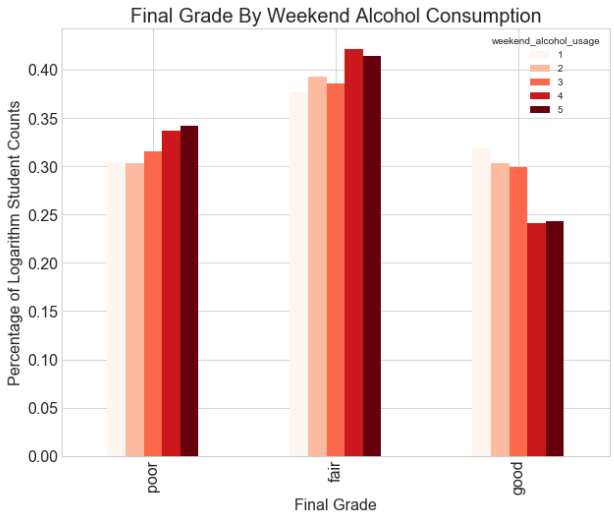

Fig. 3. Final grade by alcohol consumption

	coef	std err	t	P>|t|	[0.025	0.975]
mother_education	2.4078	0.166	14.527	0.000	2.083	2.733
father_education	1.5746	0.179	8.806	0.000	1.224	1.926

Fig. 4. Final grade by parents education level

F. Final grade by frequency of going out

The influence of hanging out more friends on students performance. Comparison between final_grade and hot attributes.

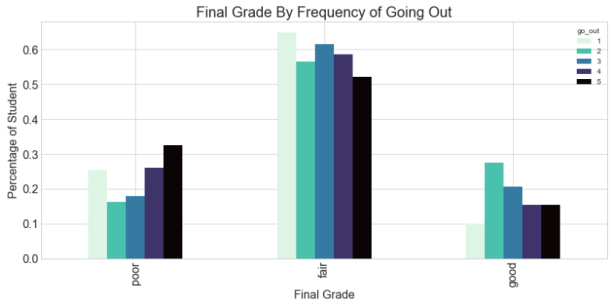

Fig. 5. Final grade by frequency of going out

G. Final grade by desire to receive higher education

Compare performance between students want to pursuit higher education and normal students. Related attributes: final_grade, desired_high_edu.

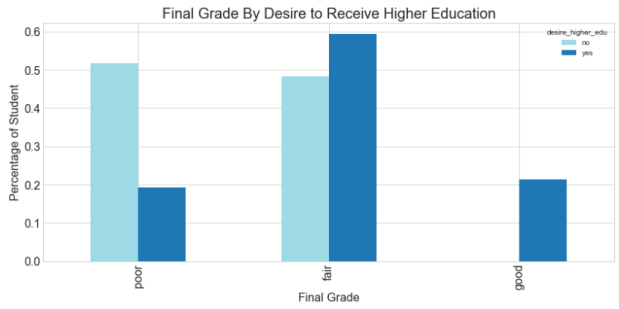

Fig. 6. Final Grade by Desire to receive higher education

H. Final grade by Living area

The effect of living area on students performance. Related attributes: final_grade, address.

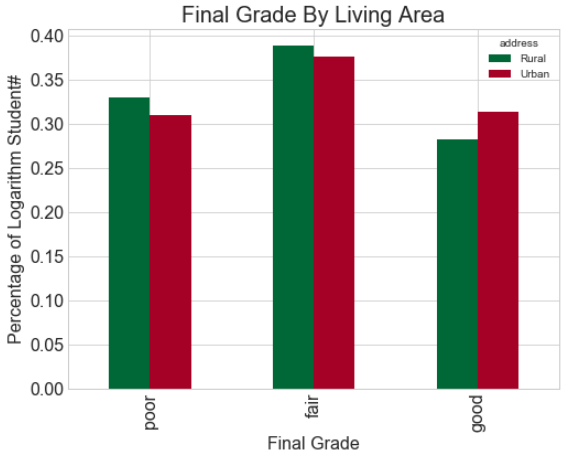

Fig. 7. Final grade by Living area

REFERENCES

[1] P. Cortez and A. Silva, "Using Data Mining to Predict Secondary School Student Performance," *Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008)*, pp. 5–12, 2008.