Proposal of Machine Learning Project:

# Alcohol Consumption and Average Grades of Students

Kewen Ding R0733305

Rui Zhu R0734658

## Motivation:

There are a lot of keys to determine the students’ academic performance, despite their attitude, interest and diligence, environment also plays a role. Nowadays an accelerating trend in alcohol use among school students has been formed. Drinking alcohol is very common for college students (Galbicsek, 2020). To what extend it positively or negatively affects students’ grades and mental health? In our research we want to find if there is any relation between students’ average grades and their alcohol consumption, family state and parents' education level.

## Dataset Description:

Link to the dataset: <https://www.kaggle.com/uciml/student-alcohol-consumption>

This dataset contains data about many attributes of the students. The inputs are

1. school
2. sex
3. age
4. address
5. family size
6. parent's cohabitation status
7. mother's education
8. father's
9. mother's job
10. father's job
11. reason to choose this school
12. student's guardian
13. home to school travel time
14. weekly study time
15. number of past class
16. extra educational
17. family educational support
18. extra paid classes within the course
19. extra-curricular activities
20. attended nursery
21. wants to take higher education
22. Internet access at home
23. with a romantic relationship
24. quality of family relationships
25. free time after school
26. going out with friends
27. workday alcohol
28. weekend alcohol consumption
29. current health
30. - number of school absences

The output of the dataset is the grade of

1. G1 - first period grade
2. G2 - second period grade
3. G3 - final grade

Seems there are many parameters of the students, we can do a lot of research based on this data.

## Method:

Students grades can be viewed as a continuous variable ranging from 0-20. It’s more reasonable to use linear regression rather than logistic regression. We can select some interesting attributes to see their relations using machine learning algorithm.

## Pointers to one relevant example of prior research on the topic:

<https://www.kaggle.com/kanncaa1/does-alcohol-affect-success>

## Reference:

Galbicsek, C. (2020). College Alcoholism and Binge Drinking - Alcohol Rehab Guide. Retrieved 11 October 2020, from https://www.alcoholrehabguide.org/resources/college-alcohol-abuse/

Feedback

先選擇一個atrribute，擬合，看看效果好不好，然後再添加一個變量，看看效果好不好，這樣一次一次的添加，效果不好就刪掉，效果好就保留，一直到10個左右

Feature selection matlab

兩種方法，互相比較一下，選擇一個最好的

Linear regression

Neural network