

# Project proposal - “Problem” - Data Mining

## Student alcohol consumption

Marco Spanò (s1045892) - Peter Horvath (s1042834)

### Problem:

Alcohol has a bad influence on human life. Despite this, alcoholic drinks are nowadays really common, especially in teenagers' life that are the ones who suffer the most from the consequences of it (the effects alcohol has on the brain and everyday's life).

The dataset we found contains records of students in secondary school that attended math and portuguese language courses with 33 attributes regarding personal info, alcohol consumption, grades and more.

Starting from this dataset we are going to make inferences such as the level of alcohol consumption of a student (based on the other columns) or predicting the grades he/she will get. Furthermore, we can also explore some characteristics of the dataset, like inspecting how alcohol influences the grades of a student.

Finally we can try to compare these results with other's taken from other similar dataset (students from different countries or older/younger) and see if there is correlation from them.

Dataset: <https://www.kaggle.com/uciml/student-alcohol-consumption#student-mat.csv>

### Our solution:

Our first aim is to start applying a decision tree classifier (and later on also a random forest) on the dataset since it is easily understandable and as well as having a good classifier we can also see which are the variables that most affect the final result.

Secondly we will apply a naive bayes classifier to compare the results we get from the decision tree to it. Also with the naive classifier we can try to inspect which attributes contribute the most to predict the outcome.

Eventually we can try to apply a small neural network on the dataset and see if it we can achieve a better accuracy, even if the dataset is not so big and the probability of it are low.

### Literature:

Paper from the researchers who shared the dataset:

[https://bradzzz.gitbooks.io/ga-seattle-dsi/dsi/dsi\\_05\\_classification\\_databases/2.1-lesson/assets/datasets/STUDENT%20ALCOHOL%20CONSUMPTION.pdf](https://bradzzz.gitbooks.io/ga-seattle-dsi/dsi/dsi_05_classification_databases/2.1-lesson/assets/datasets/STUDENT%20ALCOHOL%20CONSUMPTION.pdf)

A study on alcohol consumption of undergraduate students in the UK in 2000-2001:

<https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-8-163>

A similar study also in UK during 2008-2009:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3843305/>