# Student grade prediction

## Student Alcohol Consumption dataset

This dataset contains 33 different features, these features include things like the number of fails in classes , amount of alcohol consumed, distance to travel to get to get to school etc. A full list of the features and their meaning can be found here.
The size of the dataset is limited however (382 data entries) this will have to taken into consideration for this project.

The goal for this project is predict the final grade of a student(G3) using the available features in the dataset. The G3 is scored from 0-20 this can open up the option to predict the student's grade in binary classification (example 0-7 or 8-20)  or even projected into multi class classifications (such as 0-5,6-10, 11-15 and 16-20).

The hope is to develop a model that if given data about a student could predict the future grades of the student so that more supports could be provided to that student if the predicted grade is low.

In coding terms the features could be normalised into binary values for fields such as sex , yes or no values etc. Processes such as hyper-parameter optimization and feature selection will be used to increase the accuracy of the model and to find if there is a correlation between any individual feature and the resulting grade.

Due to the limited size of the available dataset cross fold validation will be used to assess the accuracy of the model, this would not be suitable for a real world scenario due to accuracy contamination however it should suffice for a proof of concept. Scikit-learn will be used to import and use the different available algorithms such as KNN or random forest.

## Featured Papers

Next-term student grade prediction. M. Sweeney , J. Lester ,H. Rangwala

Predicted students' grades based on free type comments data by artificial neural network. S. E.Sorour , T. Mine , K. Goda , S. Hirokawa