## Proposal

Alcohol consumption can have many short-term and long-term side effects on young people’s health and bad consequences on life and academic performance. From 1998 through 2005, 1,825 college students between the ages of 18 and 24 die from alcohol-related unintentional injuries, including motor-vehicle crashes.1 Alcohol use is also the major cause of sexual assaults on campus. Students who drink heavily are reported to have poorer performance on a test and high absence rate. 2 For many of us, adolescence is the point when we become independent person and make crucial decisions on our own. I would like to identify the factors that impact the students’ school performance, especially to check if alcohol use will affect students’ grades, with support of real-world data.

This project investigates techniques for regression problems and create models that can predict students’ academic performance. Regression is a major type of supervised machine learning (the other is classification). The goal for classification problems is to predict continuous number given a list of attributes. In this project, I will use students’ social characters and alcohol consumption as inputs to predict their academic performance measured in the final grade of Math, Portuguese, or the average of both.

## Data

The data is collected by Fabio Pagnotta and Hossain Mohammad Amran3,4 from University of Camerino. The data can be accessed through UCI Machine Learning Repository:

https://archive.ics.uci.edu/ml/datasets/STUDENT+ALCOHOL+CONSUMPTION

The data covers alcohol usage of a group of students under secondary education in Portugal along with their social, gender, study time, and school performance information. The data consists of two data-sets related to Math (with 395 observations) and Portuguese language (649 observations) classes.

1. Hingson, R.W.; Zha, W.; and Weitzman, E.R. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18–24, 1998–2005. Journal of Studies on Alcohol and Drugs (Suppl. 16):12–20, 2009. PMID: 19538908 [**http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2701090/**](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2701090/)
2. <http://www.cehd.umn.edu/fsos/projects/alcohol/consequences.asp>
3. P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.
4. Using Data Mining To Predict Secondary School Student Alcohol Consumption. Fabio Pagnotta, Hossain Mohammad Amran, Department of Computer Science,University of Camerino