Impact of Alcohol Consumption on Students

**DataNiners**

Abhishek Deshpande

Himanshu Parihar

Saranya Nakka

Suyash Nande

ITCS 6162 Knowledge Discovery in Databases

University of North Carolina Charlotte

Date: 06/08/2017

**Impact of Alcohol Consumption on Students**

This research focuses on the effect alcohol use on secondary school students’ academic performance. Past research has shown the prevalence of alcohol use among secondary school students. It has been found that most students partake in drinking alcohol. It has also been found that some students are able to successfully manage their academic work despite drinking alcohol. Despite this finding, not all students are successfully able to partake in both activities.

This topic is socially important because alcohol use is very common among students on college campuses. Alcohol use is negatively affecting some students and families by causing potential dropouts and failing classes. On the other hand, some students know how to balance their academic work with their social life and not let it affect their academic performance. The objective of this research is to discover if alcohol consumption has any predictive power or force over students grades in school.

**Dataset**

**Description and Purpose**

The dataset we chose represents the survey done over students belonging to Math and Portuguese language courses in secondary school.  It contains a lot of interesting social, gender and study information about students. Although it has many attributes to consider, our focus would be to predict the effect of Alcohol on student grades.

**Sourc**e

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.

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

**Related Work**

Going through few awesome analysis done by other data science enthusiasts from kaggale.com. We found few interesting predictions already done on this data set.

https://www.kaggle.com/calcifer/alcohol-consumption-and-average-grades

This data exploration started off by considering effect of Alcohol over grades. Unlike the expectations, the results pointed that the effect of student’s grades depended on the top two factors is the highest:  1. higher education plans 2. mother’s education but not mainly on alcohol consumption.

# Business Research/Understanding

**Project Objectives**

**Problem Domain**

 Alcohol consumption in minor students is a major issue in the society. Students start consuming alcohol way before the legal age. Here we are analyzing effect of alcohol on student’s academic performance.

**Requirements**

 To perform the analysis of alcohol consumption on students we are using the dataset surveyed from math’s and Portuguese language class namely **student-mat.csv** and **student-por.csv** and to achieve the same we will use Microsoft R language in R studio. To effectively calculate the effect of alcohol consumption on students’ performance we need to merge the two datasets. There are several (382) students that belong to both datasets. These students can be identified by searching for identical attributes that characterize each student.

**Restrictions**

Our research is limited to the survey performed on students of only secondary school. In addition to this, we have time constraint as we must complete our research within two weeks.

**Data Mining Problem Definition**

The problem that we are going to cater is that alcohol consumption is hampering the performance of secondary school students. So, the goal of the project is to discover if alcohol consumption has any predictive power or force over students grades in school. We will also check how other variables compare to alcohol consumption and weather alcohol consumption is the best predictor when forecasting a student final grades.

**Strategy**

We will use the following data mining/statistical methods to perform various data mining tasks:

* EDA (Description)
* Regression analysis techniques (Estimation and Prediction)
* K-nearest neighbor algorithm and decision trees
* Random forest algorithm