8/23/2016
Pavel Dudin
Charlotte, NC

Overview

Although we think of college as a time when young adults experiment with alcohol, the college years are rarely the first time students have faced decisions about alcohol. According to the nationally representative "Monitoring the Future Study", in 2012, 42 percent of high school seniors reported having had alcohol (more than just a few sips) within 30 days prior to the survey, and 24 percent reported binge drinking within the previous two weeks.
When considered in a developmental framework, college students face multiple, challenging transitions (Schulenberg and Maggs, 2002), including: pubertal and physical development, brain development, cognitive and moral development etc. The transition to college requires major changes in every aspect of a student's life. Students are looking for new friends who will provide support and intimacy, and they are working to develop their identity as college students (Bosari & Carey, 2001). It can be helpful to consider alcohol use and heavy drinking in relation to the developmental stages young adults are encountering (Schulenberg & Maggs, 2002).
It will be beneficial for the educational committee to know the relationships and may to prevent extensive alcohol consumption and increase overall performance in colleges.
Goals

Analyzing current database and find the relationships of Weekday and Weekends Alcohol Consumption with social factors (sex, age, parent's cohabitation status, family size, parents education, parents job), schools factors, afterschool activities and studying grades.

Predicting the overall academic performance of misuse students (3-5 on grade from 1-5) with the less drinking students (1-2 on grade from 1-5) by using the best predicting model model and validate it.

Database

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

Variables :
1 school - student's school (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira)
2 sex - student's sex (binary: "F" - female or "M" - male)
3 age - student's age (numeric: from 15 to 22)
4 address - student's home address type (binary: "U" - urban or "R" - rural)
5 famsize - family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)
6 Pstatus - parent's cohabitation status (binary: "T" - living together or "A" - apart)

7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â 5th to 9th grade, 3 â secondary education or 4 â higher education)
8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â 5th to 9th grade, 3 â secondary education or 4 â higher education)
9 Mjob - mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
10 Fjob - father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
11 reason - reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")
12 guardian - student's guardian (nominal: "mother", "father" or "other")
13 travel+time - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15 failures - number of past class failures (numeric: n if 1<=n<3, else 4)
16 schoolsup - extra educational support (binary: yes or no)
17 famsup - family educational support (binary: yes or no)
18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 activities - extra-curricular activities (binary: yes or no)
20 nursery - attended nursery school (binary: yes or no)
21 higher - wants to take higher education (binary: yes or no)
22 internet - Internet access at home (binary: yes or no)
23 romantic - with a romantic relationship (binary: yes or no)
24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29 health - current health status (numeric: from 1 - very bad to 5 - very good)
30 absences - number of school absences (numeric: from 0 to 93)

these grades are related with the course subject, Math or Language:
31 G1 - first period grade (numeric: from 0 to 20)
31 G2 - second period grade (numeric: from 0 to 20)
32 G3 - final grade (numeric: from 0 to 20, output target)

There are several students that belong to both datasets .
These students can be identified by searching for identical attributes
that characterize each student. Two .csv files will be added into the R, merged with rbind.
Unique combinations will be determined with uniquecombs()

Data analysis : Data will be analyzed and plotted. The linear, logistic regression and CVM classification and regression method will be performed . Decision what model to use will be opted depends on a errors.
Known limitations: data reflects only students from 2 schools. Different schools rate can affect the data significantly.