Factors affecting student alcohol consumption

Annie D'Souza
Electrical Engineering
Indian Institute of Technology, Bombay
20d070028

Raavi Gupta
Electrical Engineering
Indian Institute of Technology, Bombay
200070064

Sanika Padegaonkar
Electrical Engineering
Indian Institute of Technology, Bombay
20d070069

***Abstract*—With each passing day, the number of teens getting addicted to alcohol keeps increasing. This addiction not only have a significant impact on them but also affects the lives of all those who are close to them. The reasons for getting addicted to alcohol can be many, and if it isn't singled out during an early stage, the results can be catastrophic. The goal of this study is to analyse how likely it is for youth in the age group of 15-22 years to stumble into this pit by analysing and drawing conclusions from data obtained from various datasets on Kaggle. The datasets consist of various factors and the corresponding level of daily and weekend alcohol consumption. This level ranges on a scale of 1 to 5. This exploratory data analysis is followed by employing various multiclass classification models to predict the level of daily and weekend alcohol consumption of a student about whom certain attributes are known. On comparing the accuracy of these models, it is seen that a properly fine-tuned k nearest-neighbours model accomplishes the task of prediction to the best extent among all the models tried.**

I. Introduction

Students in High-School and College are exposed to a plethora of experiences, one's they've never had before. While a lot of these experiences are exciting and fulfilling, there are some that can be quite intimidating and may put a student in a tough spot, forcing them to go with the flow and act out of peer pressure. A large number of students who find it difficult to cope with various aspects of this changing phase often resort to alcohol, drugs, etc as a means of escape.

In this study, we've focused on the aspects that may cause a student to turn to alcohol and develop a drinking habit. In order to make sure that students don't rush their way into alcoholism, uninformed without knowing the consequences, it is necessary to impart general awareness to students. Apart from awareness, singling out the reasons or factors that may cause students to turn to alcoholism is also of utmost importance. This way, a student can be given the necessary guidance/ counselling before it's too late.

However, answering the question of whether or not a student is likely to turn into an alcoholic or the likeliness of a student turning into an alcoholic isn't an easy task and requires the employment of some advanced strategies or methods. This is where the field of data science and machine learning comes into play. While data science has caused a revolutionary change in various fields, it seems to have a lot of scope in the field of alcoholism prediction.

In our project, we've utilized two datasets available publicly on Kaggle and merged them into a single dataset. These datasets contain various attributes of a normal school student's life including social factors, family, scholastic environment and other general details. Among scholastic factors, corresponding to each student, there is information regarding his/her study time, failures in a subject, absences from school, higher education, etc. With regards to family, factors like mother's and father's job and education level, the presence or absence of family support in their education, etc are included. The social factors include the frequency with which a student goes out, whether or not he/she uses the internet, if he/she is involved in a romantic relationship, etc. Apart from these, factors like the student's sex, age, school, address(whether the student belongs to an urban neighbourhood or a rural neighbourhood), etc are also included. Corresponding to these factors, the level of alcohol consumption (daily and weekend) of students are given. This level is an integer ranging from 1 to 5 with 1 indicating that the student doesn't consume alcohol and 5 indicating that the student has quite high levels of alcohol intake.

We started by performing exploratory data analysis on our dataset and analysing the various attributes together. This was followed by analysing each variable separately and also its relationship with the daily and weekend alcohol consumption rates of students by plotting count plots, pie charts, bi-histograms, normalized bi-histograms and also performing the chi-square test to find out whether there is any dependency of the level of alcohol consumption on the variable being analysed.

After completing the EDA, we moved to try out various machine learning models to see which one fits the task of predicting alcohol consumption level the best. The models we tried are Random Forests, K- Nearest Neighbours, Logistic regression, Serial Vector Classification, Neural Networks and Naive Bayes. Dividing our data into training, test and validation sets, we performed grid search to optimise the hyperparameters of all the models using the validation set, used the train set to train the models and the test set to calculate the scores (here, accuracy) of each model. This process was carried out twice, once using only the relevant variables obtained using the Chi-square test and the other using all the variables of the dataset.The results were then obtained and documented.

II. Background and prior work

A. Background

Before getting into the details of this project, one must have prior knowledge of the following:

1) Chi square test of independence: This test compares two variables in a contingency table to see if they are related. In a more general sense, it tests to see whether distributions of categorical variables differ from each other. The formula for the chi-square statistic used in the chi square test is:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

The subscript "c" is the degrees of freedom. "O" is your observed value and E is your expected value.
A low value for chi-square means there is a high correlation between your two sets of data. In our case two hypothesis were taken:

- The null hypothesis H_0 : There isn't a relation between the two variables.
- The alternate hypothesis : There is an association between the two variables

In our case, a p-value was calculated, compared with the α level which was taken to be 0.05. If the p-value calculated was found to be less than α then, the null hypothesis was rejected otherwise accepted.

2) Heatmaps: A heatmap is a graphical representation of data that uses a system of color-coding to represent different values. Heat maps make it easy to visualize complex data and understand it at a glance.

One must also know about the following models that can be used to solve classification problems:

3) Random Forests: Random forests is a supervised learning algorithm which can be used for both-classification and regression.Random forest classifiers fall under the broad umbrella of ensemble-based learning methods.The key principle underlying the random forest approach comprises the construction of many "simple" decision trees in the training stage and the majority vote (mode) across them in the classification stage. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

4) K-nearest neighbours: The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems.KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression).

5) Logistic Regression:: Logistic regression, despite its name, is a classification model rather than a regression model. Logistic regression is a simple and more efficient method for binary and linear classification problems.LR is a transformation of a linear regression using the sigmoid function. The vertical axis stands for the probability for a given classification and the horizontal axis is the value of x. It assumes that the distribution of y—x is a Bernoulli distribution. The formula of LR is as follows:

$$F(x) = \frac{1}{1 + e^{-}(b_0 + b_1 x)} \tag{2}$$

Here $b_0 + b_1 x$ is similar to the linear model y = ax + b. The logistic function applies a sigmoid function to restrict the y value from a large scale to within the range 0–1.

6) Support Vector Classification:: In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

7) Neural Networks: Neural networks are complex models, which try to mimic the way the human brain develops classification rules. A neural net consists of many different layers of neurons, with each layer receiving inputs from previous layers, and passing outputs to further layers. The way each layer output becomes the input for the next layer depends on the weight given to that specific link, which depends on the cost function, and the optimizer. The neural net iterates for a predetermined number of iterations, called epochs. After each epoch, the cost function is analyzed to see where the model could be improved. The optimizing function then alters the internal mechanics of the network, such as the weights, and the biases, based on the information provided by the cost function, until the cost function is minimized.

8) Naive Bayes:: It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

B. Prior work:

As addiction of students to alcohol is a matter of great concern, there have been some papers that analyse the same. Some of them have been mentioned below:
1. Using data mining to predict secondary school student alcohol consumption: Research Paper 1
2. Is alcohol affecting higher education students performance: searching and predicting pattern using data mining algorithms: Research Paper 2

III. Data and Methodology

The data set contains the information of 1044 students and was obtained in a survey of math and portuguese students in secondary school. It covers various social, scholastic and family related aspects of the students, containing a total of 33 variables. The variables present are:

1) sex - student's sex (binary: 'F' - female or 'M' - male)
2) age - student's age (numeric: from 15 to 22)

3) address - student's home address type (binary: 'U' - urban or 'R' - rural)
4) famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
5)

Out of the 33 variables, Dalc and Walc are the features while the remaining 31 variables are the attributes. Dalc and Walc are categorical and are integers ranging from 1 to 5 with 1 representing 0 level of alcohol consumption and 5 representing the highest level of alcohol consumption.
On performing EDA on this data we notice that none of the variables contain any null values ie. each variable has a total of 1044 non null values. Furthermore, almost all variables are categorical. A good portion of the variables are integers within a certain range while the other variables are categorical and in character format eg. school, sex, address, famsize, etc. We one hot encode these variables before using them in our models. For each of the 31 attributes we performed EDA.
This EDA includes the following-

1) Count plot: To visualise the number of students against the different values an attribute takes.
2) Pie Chart: To represent the percentages of the different attribute values.
3) Bi-histograms: Represents the count of each value for the different levels of Dalc and Walc.
4) Relative bi-histogram: Represents the percentage of each value of the attribute against the levels of Dalc and Walc.
5) Box and whiskers plots: To visualise the median, quartiles and outliers.
6) Chi-square test for association: here we calculate the p value of the attribute and compare it with the significance level. If the p value is lesser than the significance level, the null hypothesis is neglected and it is inferred that the attribute and Dalc/ Walc are dependent. However, if the p value is greater than the null hypothesis, the attribute and Dalc/ Walcare are declared as independent.

Thus, using the chi square test, we see that both Dalc and walc depend upon the following variables:

- Sex
- Age
- Traveltime
- Studytime
- Freetime
- goout

Whereas, only Dalc depends upon the following variables:

- Reason
- Failures
- Higher
- Famrel
- Health
- Absences

Also, only Walc depends upon the following variables:

- Fjob
- Schoolsup

To get a comprehensive idea about the correlation of the individual columns with daily and weekend alcoholism, heat plots were created.
The darker the color of a particular variable, the more it is related with the respective type of alcoholism.

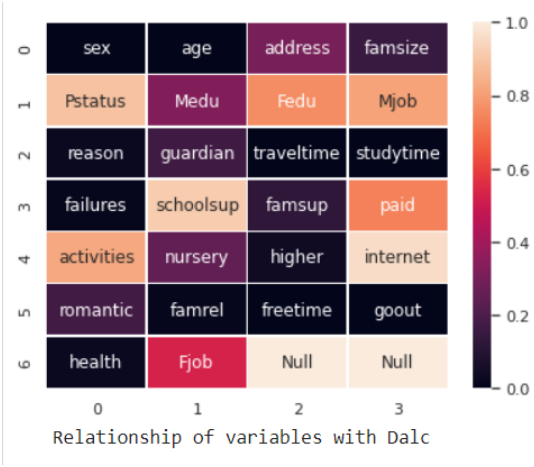


Fig. 1. Using the p-value obtained for each variable heat plot was made

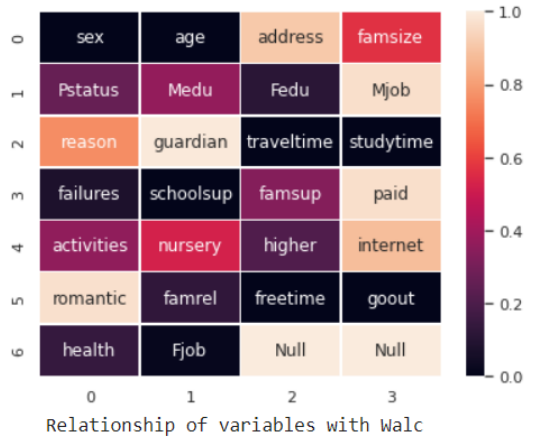


Fig. 2. Using the p-value obtained for each variable heat plot was made

Now that we have realised which variables are more relevant for our task, we take a deeper look into them-

A. *Sex*

The gender of a person is often considered to be a strong factor while analysing alcohol consumption. The general consensus of society is that men are more prone to alcoholism than women. But is this true in our case? To check this, let's look at the data collected. From the bar graph and pie chart, it is visible that the data contains a higher proportion of female students than male. From the histograms, male students show higher absolute counts of higher levels of weekday and weekend alcohol consumption. On examining the relative plots, we observe that here too, male students show higher levels of weekday and weekend alcohol consumption.

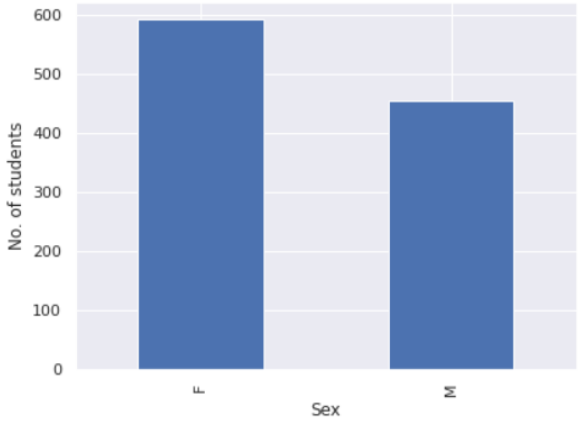


Fig. 3. Distribution of students according to their gender

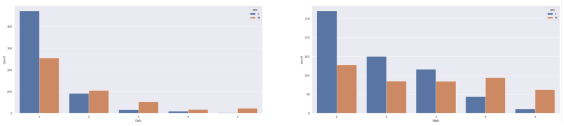


Fig. 4. Comparison of students in absolute terms according to their gender

On computing the chi-square statistic and p-value for the

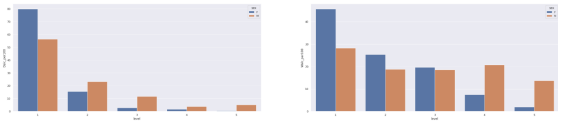


Fig. 5. Comparison of students in relative terms according to their gender

hypothesis test of independence, we see that weekday and weekend alcohol consumption have a high dependency on sex.

B. Age

Age plays a crucial role in studying alcohol consumption. Younger population might show lower levels of alcohol consumption due to restricted access to alcohol. Let's look at the role of age of the students in our dataset. From the piechart, it is visible that the data contains the highest proportion of 16 and 17 year olds, followed by 18,15 and 19 year olds. The data contains a relatively lower proportion of students aged 20 22 years. Since our data mostly contains students below 18 they have a higher absolute count of the levels of alcoholism but as it can be seen from the second pair of bar plots, older students(19-22) have higher relative counts of higher levels of weekday and weekend alcohol consumption. Upon computing the chi-square statistic and p-value for the hypothesis test of independence, it is observed that the p-value is less than α therefore we can reject the null hypothesis, and conclude that alcoholism and age levels are associated with each other.

C. Father's Job

Our dataset mainly contains the following student's fathers jobs-

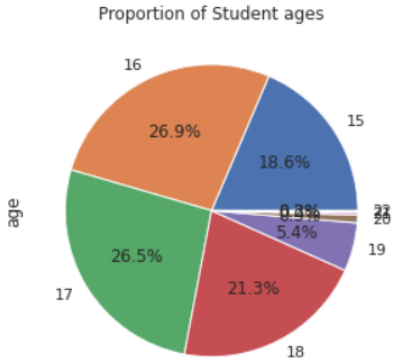


Fig. 6. Distribution of students in the data according to age

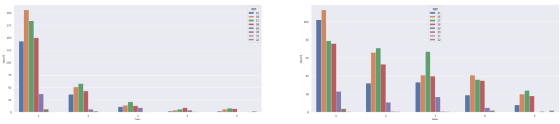


Fig. 7. Absolute consumption of alcohol by students according to their ages

- Healthcare: 3.9%
- Stay at home:5.9%
- Teacher: 6.2%
- Services: 28%
- Other jobs: 55.9%

From the data, we observe that the fathers of most of the students have jobs other than the ones mentioned. Consequently, it is expected that these students have higher numbers in almost all levels of alcohol consumption. Comparing the trends between daily and weekend alcoholisms here, it can be

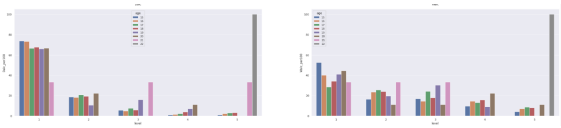


Fig. 8. Relative consumption of alcohol by students according to their ages

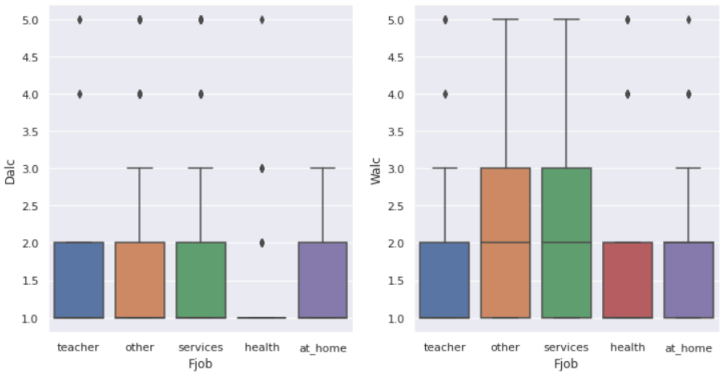


Fig. 9. Daily and weekend alcoholism box-plots

seen that students whose fathers are in services or have other jobs tend to consume more alcohol in weekend. Moreover, students whose father work in health care don't have tendency to consume alcohol on a daily basis.

D. Reason(for choosing the school):

The following reasons are observed in our dataset:close to 'home', 'school', 'reputation', 'course' preference or 'other' .Reason may affect alcohol consumption. For eg: if the school was chosen because of its reputation or its courses but it's far away from a student's home, this particular student might find ways to escape the supervision of his parents and teachers and consume alcohol or engage in other such activities.
Let's see if this theory holds for our dataset. As can be seen,

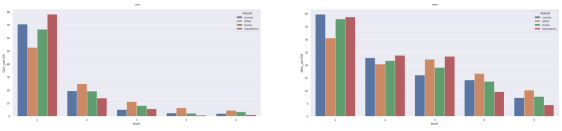


Fig. 10. Relative alcoholism according to reasons for being in the particular school

relative levels of all 4 classes remains the same in Daily alcoholism and weekend alcoholism with students stating their reason for joining the school as 'others' as the highest.

E. Traveltime

The data was divided into 4 parts on the basis of the traveltime of students to and from their home to school with the following demarcations:

- 1:less than 15 minutes
- 2:15 to 30 minutes
- 3:30 min to 1 hour
- 4:greater than 1 hour

Although it doesn't seem very intuitive, travel time can affect alcohol consumption. Let's look at our dataset for example. It can be observed that most of the students in the dataset have travel time less than 15 minutes. Consequently their absolute

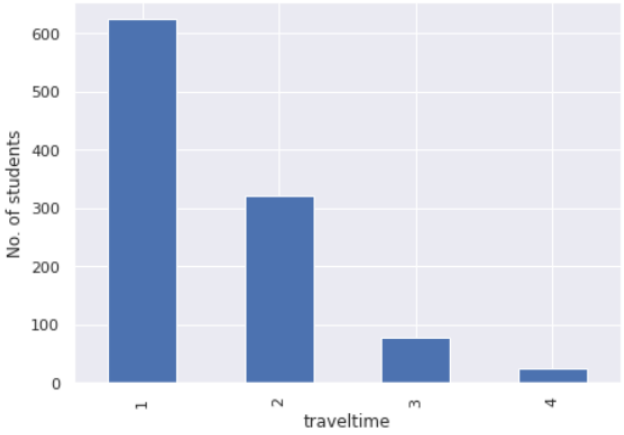


Fig. 11. Bar plot for traveltime

alcoholism is higher but upon comparing percent alcoholism a drastic increase is seen upon those who have traveltime greater than 1 hour have a significant increase in their level of alcoholism.
Upon studying the results of the chi² test too it can be concluded that both daily alcoholism and weekend alcoholism have a strong correlation with travel time.

F. Studytime

The amount of time a student spends studying directly impacts the activities that they engage in.The data contains a higher number of students having weekly study time between 2 to 5 hours.
Despite the fact that they have higher absolute numbers, students having study time even less have higher levels of alcohol consumption. The gap somewhat smoothes out in the percent bar graph but group 1 still maintains the lead in both daily alcoholism and weekly alcohol consumption.

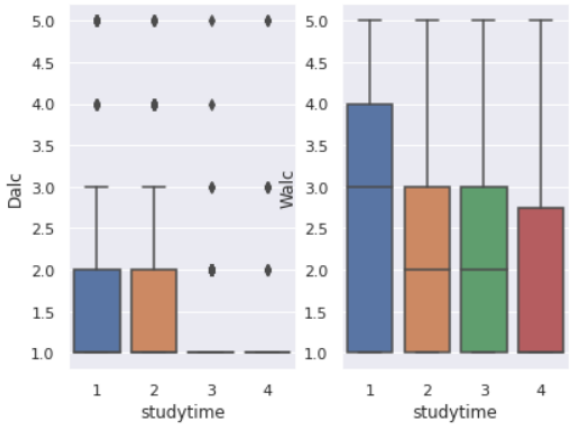


Fig. 12. Box plots for studytime

From the box and whiskers plot it can be observed that group 1 have the highest median in weekend alcoholism followed by 2, 3 and 4.This is intuitively correct too. Since the number of students are very less in category 3 and 4 in Dalc most of them are treated as outliers.

G. Failures

Failures can have a deep impact on the mental health of students. They might demotivate and depress them and the students might develop destructive tendencies like alcohol and drug abuse. The dataset is highly dominant towards kids which don't have prior class failures therefore the plots for absolute daily and weekend alcohol consumption also favour class 0.Upon observing the percent bar plots, it can be seen on a relative term that people having a higher number of past failures are more prone to alcoholism.

H. Extra school support for education

We expect that schools should be supportive of the extra education of students. Still, for most of the students in the dataset, additional support for education was not provided by

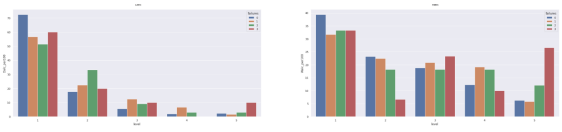
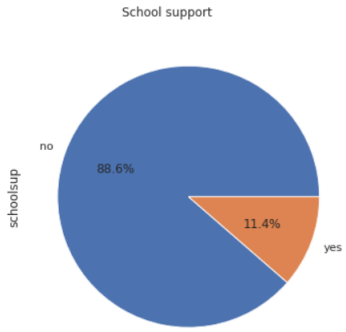


Fig. 13. Analysis of how failures relates with alcoholism



the school, as can be seen in the pie chart. Upon observing the p-value, since it is more than 0.05 for Daily alcoholism, we cannot reject the Null hypothesis, but for weekend alcoholism, the p-value is, in fact, less than 0.05. Therefore, there exists an interdependence between school support and weekend alcoholism but not between daily school support and daily alcoholism

I. Higher education after school

Our dataset overwhelmingly contains students who want to pursue higher education, therefore in absolute terms these students have majority in all levels of alcohol consumption. Relatively, as the level of alcohol consumption increases people who aren't interested in pursuing higher education tend to have higher alcohol consumption levels.

Weekday alcohol consumption has a significant amount of interest in whether the student wants to pursue higher education or not but weekend alcoholism doesn't.

Dalc	1	2	3	4	5	Total
higher						
no	46	27	7	3	6	89
yes	681	169	62	23	20	955
Total	727	196	69	26	26	1044

Fig. 14. Contingency table between higher education and Dalc

Walc	1	2	3	4	5	Total
higher						
no	29	15	17	14	14	89
yes	369	220	183	124	59	955
Total	398	235	200	138	73	1044

Fig. 15. Contingency table between higher education and Walc

J. Quality of family relationships

Family is the foremost example of how a student perceives the outer world. If the relationship is more strained there, it can be a major reason for the student to resort alcoholism. Here 1 corresponds to very bad relationship while 5 to an excellent one. Checking the relative barplots, it is clear that students that have a unhealthy relationship with their family are more likely to fall into the pit and must be helped as soon as possible.

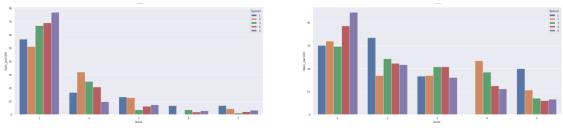
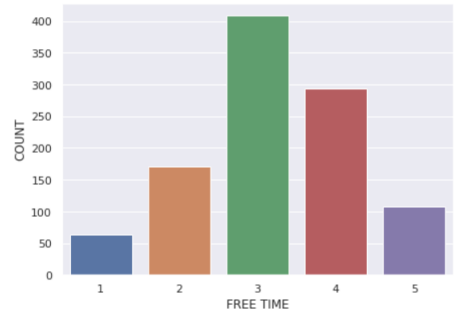


Fig. 16. Levels of alcohol consumption v/s family ties

K. Freetime

It is of course expected that the more the freetime a student has, the more his/her mind will off and seek solace in the form of alcohol. Most of the students in our dataset had intermediate amount of freetime. Despite of the fact that students having a high level of freetime have less representation in the entire data, they show the highest level of daily alcoholism in absolute numbers!



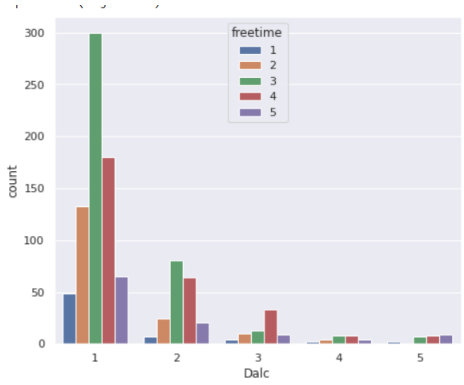
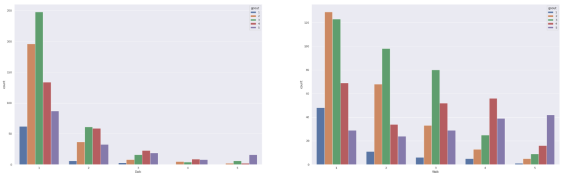


Fig. 17. Despite of having the lowest numbers, students having high freetime have more alcoholism rates

L. Frequency of going out with friends

It is a common perception that if a teenager ventures out with friends more he/she is more prone to drinking alcohol. Upon taking a closer look at our results, it turns out that this is indeed true. Our dataset contains levels of going out with 1 being very low to 5 being very high.The majority of the students belong to level 3 followed by 2, 4,5 and 1.
Despite the fact that level 5 contributes less number of students to the data set, they have the highest level of alcohol consumption both in absolute and relative terms as can be seen in bar plots.

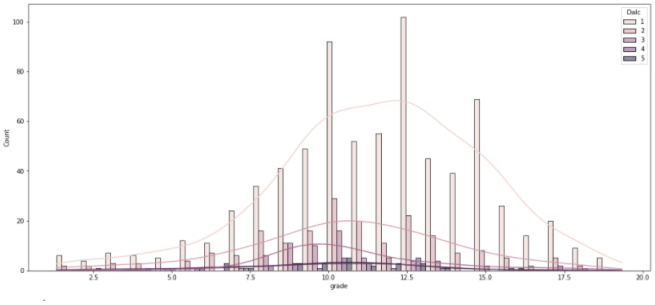


M. Health

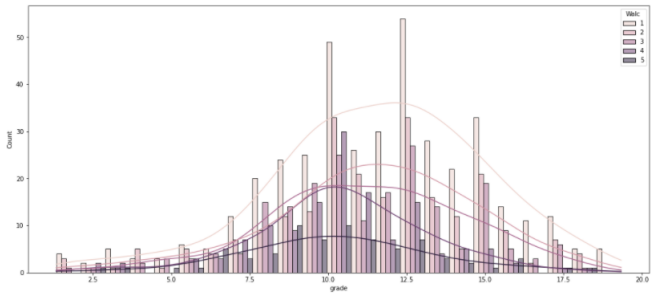
Majority of the students in our dataset have moderate to very good health status(3-5).
Since our data mostly contains students with moderate to very good health status (3-5), they have higher absolute counts of the levels of both weekday and weekend alcoholisim. We do not observe any clear trend in relative levels weekday and weekend alcohol consumption from the relative bar plots.
On computing the chi-square statistic and p-value for the hypothesis test of independence, we observe an intriguing result-weekday alcohol consumption depends on the health status of students but weekend alcohol consumption does not.

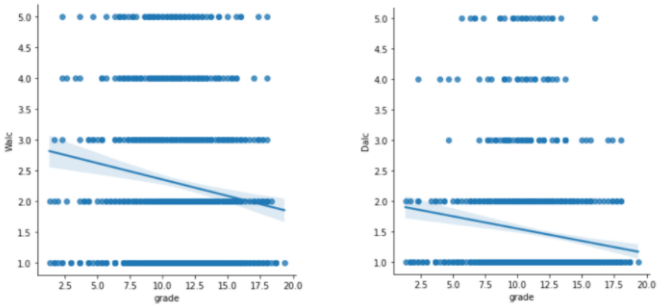
N. Grades

The original dataset consisted of grades of students in the three periods of their year. To reduce redundancy the three columns were merged into a single column with the values of the new one being the average of the previous three. Next histograms were plotted with hues taken as Dalc and Walc to visualise the relationship between the two. The plots obtained were as follows -
From the kde plots, we can see that as the level of alcoholism



rises the peak of the distribution shifts leftwards i.e towards a lower grade in case of both daily and weekend alcoholism.
Linear model plots were also plotted to check the dependency





of grades on the student's alcoholism levels. As observed from the previous graphs, this confirms that Dalc/Walc have an inverse relationship with the frequency that students venture out with their friends with Walc being slightly higher than Dalc.

IV. Experiments and Results

Having a cleaned up data now with fewer variables, 6 supervised Machine learning models were employed on the variables to predict the level of alcoholism among completely new students. The models were as follows-

A. Logistic Regression:

The model was trained twice-once using all available variables and the second time using only the variables on which Dalc/Walc dependent. These variables were found using the results from the chi square test performed in the EDA. The following accuracy scores were observed:

Performance of Logistic regression			
	Using all variables	Using only dependent variables	Improvement
Dalc	0.678	0.694	0.016
Walc	0.529	0.567	0.038

Thus, we get a slight improvement in our accuracy scores when we use only dependent variables. However, the accuracy isn't too high and thus, logistic regression cannot be considered as a highly reliable model.

B. Support Vector Classification (SVM for Classification):

The model was trained twice-once using all available variables and the second time using only the variables on which Dalc/Walc dependent. These variables were found

using the results from the chi square test performed in the EDA. The following accuracy scores were observed:

Performance of SVM			
	Using all variables	Using only dependent variables	Improvement
Dalc	0.682	0.686	0.006
Walc	0.464	0.579	0.115

Thus, we get a slight improvement in our accuracy scores when we use only dependent variables. The improvement in Walc was more significant than the improvement in Dalc. However, the accuracy isn't too high and thus, support vector classification cannot be considered as a highly reliable model either.

C. Random Forest

To find the optimal hyperparameters, a grid was formed and a random search was employed to find the range of optimal hyperparameters, after that a new grid was formed a grid search was performed so that the final hyperparameters obtained were the most optimal ones. The model was validated using a k-fold search and final accuracy of 68% was obtained in the estimation of Dalc while an accuracy of 53.99 % was obtained in the estimation of Walc.

D. K-nearest neighbours

The same method as of Random forest was employed in estimating the parameters for this model. Among all the models used, this model gives the highest accuracy of 83.78% in the prediction of Dalc. However, the accuracy for Walc remained at 64.54%.

E. Neural Network

We used tensorflow and keras to implement this model. The data we had was divided into train, test and validation sets. The validation set was used to find the optimal values of the hyperparameters like the units in dense and the learning rate. This random search was performed for a 3 layered neural network. While predicting Dalc, the scores for them are as follows:
3 layers: While for Walc, the optimal hyperparameters are

Units	64	64	16	16	8
Learning Rate	0.1	0.001	0.1	0.01	0.01
Score	0.6921850244204203	0.6842105388641357	0.6794258356094366	0.6794258356094366	0.6794258356094366

learning rate = 0.1, units = 32, and the accuracy is 40.19%

F. Naive Bayes

The dataset was divided into test and train sets. Grid search CV was used to find the optimum value of the hyperparameter '$var_smoothing'$ which is a stability calculation to widen (or smooth) the curve and therefore account for more samples that are further away from the distribution mean. The optimal value of the hyperparameter was: '$var_smoothing' = 1.0$ which gave an accuracy of 68.421% for Dalc and '$var_smoothing' = 0.0284$which gave an accuracy of 38.516% for Walc.

V. Learning, Conclusions, and Future Work

In conclusion, we can predict the weekday alcohol consumption of students with an accuracy of 83.78% and weekend alcohol consumption with an accuracy of 64.54%. Among the various models used, k-nearest neighbour model is best suitable for this task.
This project enhanced our knowledge about various supervised machine learning algorithms. We also learnt to work as a team, divide work, manage time and meet deadlines. In the future, we plan to finetune our EDA and find ways to make our models more accurate so as to obtain a greater accuracy.

VI. Contribution of team members

Each of us performed EDA on roughly 9-11 variables each. We used a total of 6 models in our project which were distributed amongst us as follows:
Annie:Neural Networks and Naive Bayes
Raavi: Random Forests and K nearest neighbours
Sanika:Logistic Regression and Support Vector Classification
Abstract,Introduction and overview parts of the report were handled by Annie; Background and prior work by Sanika(Background) and Annie(Prior work);Data and Methodology part was assigned to Raavi and Sanika; the Experiments and Results part was divided equally amongst the three of us; Learning, conclusions and future work was handled by Raavi. Raavi converted the final report into a Latex document. All three of us contributed to the video.

Acknowledgement

We would like to thank our professors, Prof. Amit Sethi, Prof. Manjesh K. Hanawal, Prof. Sunita Sarawagi and Prof. S. Sudarshan for teaching us this course, without which we would not have been able to complete this project. We would also like to thank all the TA's involved with this course, who helped us a lot in doing the various assignments, and completing this course.

References

[1] "Dataset":https://www.kaggle.com/uciml/student-alcohol-consumption
[2] "Definitions":https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/p-value/,https://www.optimizely.com/optimization-glossary/heatmap/,https://www.sciencedirect.com/topics/computer-science/random-forest-classifier,https://www.javatpoint.com/machine-learning-random-forest-algorithm
[3] https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761
[4] https://www.sciencedirect.com/topics/computer-science/logistic-regression
[5] https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/
[6] https://towardsdatascience.com/classification-using-neural-networks-b8e98f3a904f
[7] https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/
[8] https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c