Team 10

Table of Contents

Tables

Figures

A Data-driven Framework to Identify and Asses Risk Factors Leading to Suicide Ideation and Attempts Among Youths

Dave, Shivam[1]; Gohil, Hitesh[1]; Patel, Aditya[1]; and Mukherjee, Sayanti[1,2]

[1] University at Buffalo, The State University of New York, Buffalo NY, United States

[2] sayantim@buffalo.edu (corresponding author email)

Abstract

Suicide is one of the leading causes of death in United States. It is not only seen in adults but also among the youths attempting such kind of risk. In this research we propose to develop a data-driven framework to identify and assess the impact of the key risk factors leading to an increased rate of suicide ideation/attempts, leveraging advanced statistical learning techniques. We implemented our analysis using the nation-level data provided by the Youth Risk Behavioral Surveillance System (YRBS) from the year 1991 - 2017. Our research will present analysis of the various risk factors that are found to significantly impact the consideration of suicide attempts among the youths. Implementing supervised learning techniques, namely General Linear Models and decision trees. We found that for a particular dataset (overall data includes attempted and not attempted suicide, data classified based on gender), different methods provided the best model fit based on in sample and out of sample error. The outcomes of this research will provide a deeper insight and knowledge about the complex interplay of the key risk factors that leads to frequent suicide ideation/attempts in the younger generation; thus, help in informed decision making and development of effective prevention strategies for the youth suicides.

Keywords: suicide ideation/attempts; risk factors; youths; predictive analytics, statistical learning theory

Introduction

Suicide ideations and attempts are abominable for any country; in fact, it is one of the leading causes of death in the United States [1,3,4,17,18,19]. Although committing suicides are more often observed among older than the younger generations, still it is one of the primary causes of death in late childhood and adolescence globally [2,4]. Such acts of committing suicides not only result in direct loss of many young lives, but also has disruptive psychosocial and adverse socio-economic effects [17]. Moreover, it has been observed that the suicide rates are increasing since the past 20 years [13,20]. For both male and female, the factors involved in attempting suicide may vary as there are more female suicide attempts and suicide rates than male in youths [17,20]. The data used for our study is the YRBS survey data of 9th – 12th grade school from different states who had the participation rate more than 60% every year. Thus, there is a need of analyzing the trend of the suicide attempts among the youth and investigate the various risk factors impacting suicidal behavior in the youths.

Literature Review

It is difficult to find the data of suicide attempts as it is very sensitive topic. Therefore, researches have made their own data by taking surveys in schools and high school of the United States [1,4,7,9]. There are many communities who have taken an initiative to conduct a survey by their own in schools by providing their own questionnaire in accordance with youth risk factor behavior survey system. More than 13 nations have analyzed the relation between alcohol consumption and suicide rate where in 10 out of 13 nations were positively associated [4]. The overall effect of alcohol causing death is more in youths for both the gender [1-3,7,8,15]. Alcohol policy also helps in preventing the suicide and from the study of 1976-1999 data, male suicides have been reduced but there is no impact in female [16]. There are many researches done to find out which category of population is majorly affected and the reason behind this attempt [1 – 5, 7 - 9]. Factors such as living with single parent, feeling of depression, alcohol, previous suicide attempt, aggressive behaviors are important risk factors for youth [3,7,13]. In late 2000, researchers have found the suicidal behavior among adolescent patients were affected by alcohol abuse, past suicidal behavior, depression and aggressive behavior [5]. In 2005, 25.6% lower rates of early alcohol use were observed according to the survey of Youth Risk Behavior Survey (YRBS) compared with 1991 (32.7%) [13]. Alcohol consumption has been one of the important factors for adverse health which results in unintentional injury [2]. In different psychiatric condition, impulsivity is related to suicide and self-destruction behavior [4]. Alcohol and psychiatric disorder have a relationship which results in suicide risk [4,14].

In Ireland, young male mortality rate rapidly increases with high alcohol consumption as compared to suicide rate, unemployment rate and alcohol consumption [11]. From Signs of Suicide (SOS) data, planning and unplanned suicide has a key difference with usage of alcohol resulted in youths who does not report suicide ideation, drinking alcohol while down increases the risk of self-reported attempts [7].

It is found that alcohol use disorder (AUD) increases the risk of suicide ideation, attempts and completion which did not include any age, gender, race as a significant factor [14]. From the YRBS 2004-2005 data which included United States preteen, teenagers, adolescent and adult data found that children who starts the alcohol at their preteen age are more likely to have suicidal attempts compared with non-drinkers who are more violent in behavior, sexual assault, fighting, weapon carrying [2,3]. It shows that these effects are strongly shown by girls which makes them to attempt suicide or getting the thought suicide as compared to boys [2]. Other than demographic, depression, antisocial behavior, researchers have also found that due to insufficient amount of sleep causes them to have suicide ideation. Example in South Korea adolescents encounter with highly competitive education environments and college entrance exam which affects the rate of suicide ideation [8]. Major predictors can be said to be depression, stress, less sleep, alcohol and other substances use, leads to suicide attempts and suicide ideation among youths [1,3,4,7,9,11,18,19].

In various parts of the world, researches have also taken weather as a factor which may have an impact on suicide [6, 10]. This will help the state government to take measurable action against the alcohol by generating policy or increasing the rate so that it will reduce the consumption of alcohol in the that particular state. Suicide taking place due to hanging, drowning and jumping was compared with the metrological factors and concluded that drowning was the one which was associated with meteorological parameters in men [6]. Stepwise regression weather model was developed adjusted for season [10]. It is concluded that suicide rates are higher associated with high temperature but the relative risk is low and season played an important factor in younger population [10].

The methods used for finding the significant predictors were logistic regression, multinomial logistic regression, stepwise regression etc. [2,3,7,8,5,12]. Other than linear model, we want to implement decision tree methods to see what difference it gives in building predictive model for assessing the risk factor involved in suicide attempts. Our aim of the study is to utilize the whole YRBS data and predict the risk factors leading to suicide which has more predictors than the previous studies such as type of drug use, use of cigarette, sexual intercourse, bullying, weapon carrying etc. This will have a better understanding of why the youth takes such a heavy risk. Further our study can be classified for different gender, male and female. Since the female are experiencing more suicide attempts and completion than males, there might be different factors from male leading to take such a step.

Data Collection, Preprocessing and Analysis

The Youth Risk Behavior Surveillance System (YRBSS) monitors six categories of priority health behaviors among youth and young adult. All the predictors in the dataset are categorical variable and since our topic is align towards alcohol and suicide, therefore we decided to visualize the alcohol and suicide related questions. The data collected by filling out the questionnaire provided by YRBSS. Table no.1 shows us the data is collected after every alternate year. The people who had suicide ideation is “Yes” and who didn’t have is “No”. Some of the people didn’t answer the question. From the plot in Fig.1 can be interpreted as increase in suicide ideation more than 10 years.

Year	**2003**	**2005**	**2007**	**2009**	**2011**	**2013**	**2015**	**2017**
No	12705	11481	11767	13871	12869	11232	12626	11982
Yes	2453	2341	2092	2349	2424	2259	2808	2571

Table 1 Frequencies of Suicide Ideation over 20 Years

Survey data is quite challenging to handle, meaning for predictive model building we need to analyze all the variables which are in discrete ordinal structure. The dataset was obtained from CDC we in access format and using ‘RODB’ package we obtained the dataset into R. Going through the variables which were selected as predictors were found from the literature review as well as from the questionnaire of the survey data. We performed *chi-squared* test the data to identify the possible relations between the response. Since the features are not causal other than 3 variables which are directly related to ‘Number of attempts’ like consideration, planning and feeling sad or hopeless. The survey response had 150 plus variables. After removing the variables with over 70% NA values and cleaning of data, the data obtained was around 195,000 observations from the survey data conducted from the year 1999 to 2017.

From the *'goodman and krushkal tau's correlation test'* as shown in figure, we can observe that intense responses (where $K = 1 < 2 < 3 < 4 < 5 < 6 < 7 < 8$) like heavy smoking, higher occurrence of physical fighting, more use marijuana on daily basis are linked to suicide attempts. The basic goal to analyze the data was to identify important behaviors that were impacting the ideations and attempts overall so the '$Q28 - No. of\ attempted\ suicides$' was converted into 2 classes according to responses as attempted and not attempted. Further, the data with response 'Yes' was categorized into 2 classes to analyze the severity of attempts i.e. what factors are further causing a youth to attempt suicide more than once. Implementing below mentioned methods on training and test set was forcing the method to learn on negative responses. The approach of down-sampling the data resulted in interesting revelations that some features, even though not causal can cause a significant impact on suicide ideations and attempts. To get even more insights on specific reasons we further categorized the data into responses from male and female. The 33 Predictors obtained after pre-processing are as follows:

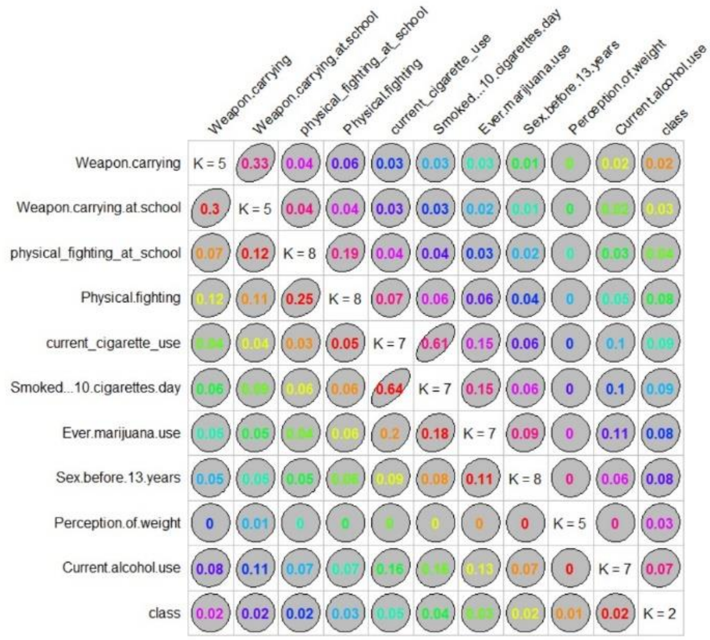

Figure 1 Correlation Matrix- Krushkal Tau Test

Table 2 Selected Predictors

Predictors	
State	Number of Suicide Attempts in past 12 months
Year	Ever Cigarette Use
Age	Initiation of tobacco products
Grade	Initiation of Alcohol
Race	Initiation of Marijuana
Gender	Use of birth control Pills
Using seatbelt while driving	Using drugs before sexual activity
Carrying weapons on and off campus	Perception of weight (How do you think about your weight?)
Using drugs on campus	Current use of tobacco products
Being Threatened at school	Current use of Alcohol
Having sex before the age of 13	Current use of Marijuana
Safety Concerns at school	Ever Alcohol Use
Physical Fighting on and off campus	Ever Cocaine use
Feeling hopeless in past 12 months(causal)	
Considering Suicide (causal)	
Planning a Suicide (causal)	

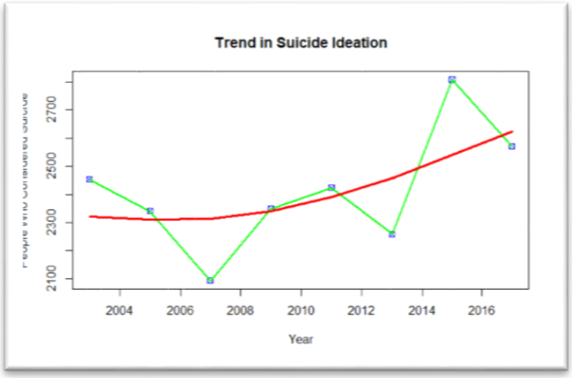

Figure 2 Trend in Suicide Ideation

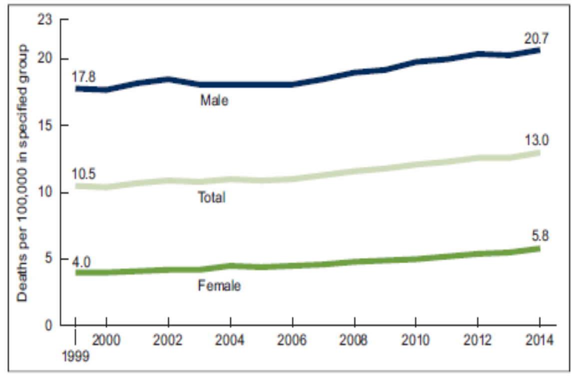

Figure 3 Trend in Deaths from Suicides

From the Figure 2 Trend in Suicide Ideation it can be clearly seen that number of suicide attempt made by youth is increasing over the years. Also, there is a bit increase in the numbers of both male and female in United States [20]. It was pretty much clear to assess the risk factors in both male and female which results in increase of suicide attempts every year.

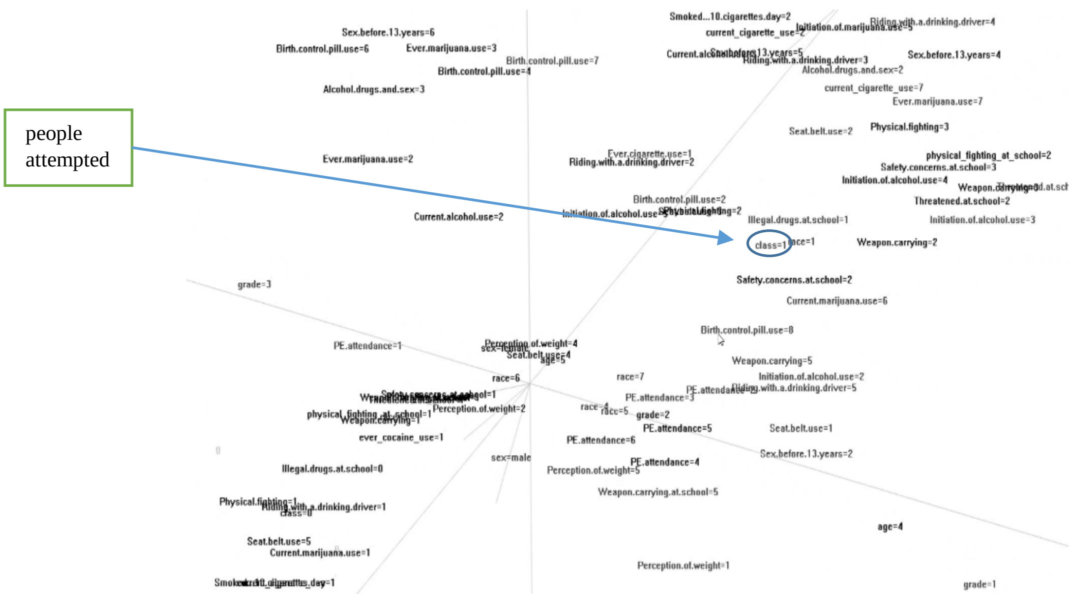

Figure 4 Principal Component Plot

Principal component analysis (PCA) was done to reduce the dimension of the variable according to class 0 representing ‘No suicide attempt’ and class 1 representing ’suicide attempt made’. It was difficult to visualize in two dimension, Figure 4, therefore it was analyzed in 3 dimension.

After analyzing the predictor and finalizing the Reponses the following figure shows our implementation of data into different models.

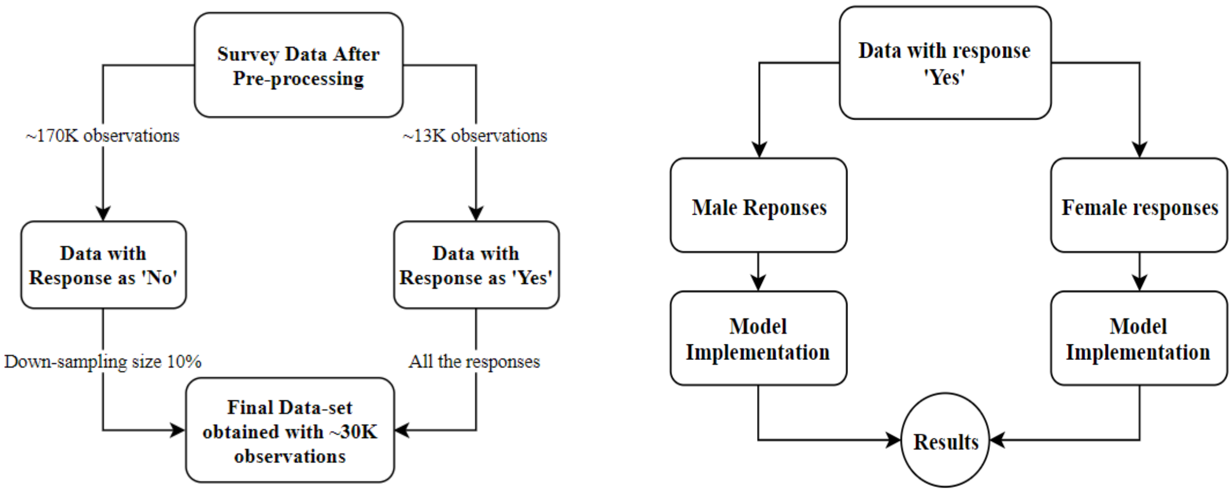

Figure 5 Methodology to Pre-process data

Methodology

Implementing the aforementioned predictors while analyzing the data we found that casual features, considering suicide, feeling sad and hopeless making a suicide plan. This helps us to finalize predictors for the response from the input data and make inferences for response variable. Thus removing the casual variables helped shed more light on classifying other predictors with the response class. We trained our data using non parametric method. The method does not have any assumption about underlying function. It is better than parametric method because we do not need to worry about right features. They result in higher performance models for prediction. However, the ease of interpretability will reduce. But for the better interpretability we implement various classification methods to understand the variable importance and its relationship with the response. The various methods which we are going to implement is Generalized linear models(GLM), Principal Component Analysis, Recursive Partitioning, Classification Trees, Bagging, Random Forest, Support Vector Machines [21]. We want to see how the single tree differs from ensemble methods for the categorical and frequency data. Advantage of ensemble tree is they are more robust and are not sensitive to a small change in input. They are used for reducing the variance of a statistical learning method and have better predictive performance than single tree methods. Brief review of the above methods is described as follow:

a.) Generalized Linear Model (GLM)

GLM is the extended version of Linear Regression model in which the variables are not normally distributed. They are termed as general because they can take variables with distribution such as Bernoulli and Gaussian. The observation 'μ' ϵ {0,1} following Bernoulli Distribution with mean parameter conditional on predictor value. Therefore, the error term in logistic regression does not exist.

$$logit(p) = \log\left(\frac{p(y=1)}{1-(p=1)}\right) = \beta_0 + \beta_1 x_{12} + \cdots + \beta_p x_{in}\ for\ i = 1,2,\ldots,n$$

Where μ is the link function which links the expected values of the response to linear combination of predictors. This means that constant change in predictor results in change of response variable. β's are the unknown parameters.

b.) Classification Trees – Recursive Partitions

It is similar to Regression Tree, expect it is used for qualitative response. We can predict each observation belongs to most commonly occurring class of training observation in the region to which it belongs. In this method recursive binary splitting is used to grow the tree. For making these splits, instead of RSS, classification error rate is used. Meaning the fraction of training observations in that region that do not belong to most common class.

$$E = 1 - \max(\hat{p}mk).$$

where $\hat{p}mk$ represents the proportion of training observations in the m^{th} region that are from k^{th} class. Other method is entropy which takes on a value near zero if $\hat{p}mk$ are near to zero or one.

$$D = - \sum_{k=1}^{k} \log(\hat{p}mk) * \hat{p}mk$$

For mth pure node, entropy will take a small value. This is used for pruning the tree whereas classification error rate is used for prediction accuracy for the final pruned tree. The advantage of tree model over linear model is that they can be interpreted easily and can be displayed graphically. Also handling qualitative predictors is easy as there is no need to create dummy variables.

c.) Bagging –

Bootstrapping or Bagging is a general procedure for reducing the variance of a statistical learning method. With a given set of observations $Z_1 \ldots Z_n$ each with variance σ^2 the variance of mean $\bar{Z}$ can be given as $\frac{\sigma^2}{n}$. Averaging a set of observation reduces variance. We take repeated samples sets from a training dataset and define a separate prediction model using each set B and averaging them as

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}^{b}(x)$$

d.) Random Forest –

The major difference between Bagging and Random Forest is that, a new sample of *m* predictors is taken at each split and we choose m = $\sqrt{p}$, meaning predictors taken at each split is equal to square root of total number of predictors. This method helps in reducing the variance of a single tree by forcing each split to consider only a subset of predictors. This can be thought as "decorrelation" of the trees, making average of the resulting trees more reliable and less variable. If B (bootstrapped regression trees) increases, the method will not over fit. Therefore, we can use large value of B for error rate to settle down.

Random Forest searches for the best feature among the random subset of features rather than most important feature for splitting a node. This adds the additional randomness to the model while growing trees. It builds multiple decision trees and sums up together to get an accurate and stable prediction.

e.) Support Vector Machines -

The SVM is a generalization of a simple and intuitive classifier called as maximum margin classifier. It classifies response based on a hyperplane which is a subspace for dimensions/predictors (p-1). The mathematical definition of hyperplane is

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots = 0 \; for \; n \; dimensions/predictors.$$

The above equation defines the hyperplane and the data points $X_1 \ldots X_n$, for n dimensions a hyperplane classifies the data into two halves and helps algorithm to identify which data point lies on which side accordingly. The maximum margin classifier decides a hyperplane by maximizing the distance of the hyperplane to margins and setting the margins closest to the classifications of data points and the points the help decide a margin are called support vectors.

f.) Model Accuracy -

To measure the performance of a model depends upon how well it predicts on an independent test sample. Minimizing the generalized error, the bias variance trade-off is the key. By this, we mean that the model should not be very sensitive to the test data and should not have more accuracy. To achieve this, we do cross validation for training the model. K-fold cross validation is used to estimate the prediction accuracy. The algorithm of this method is, it first splits the data into equal 'k' number of folds with equal number of observation in each fold. The model is fitted to k-1 subset and the kth fold is used for predicting the accuracy of the model. In our project we have split the data into 80% - 20% as training and test data respectively using 10 – folds. Since we are implementing models on ordinal data, balanced accuracy is a measure of predictions given by the model, identified as false positive and true positive.

		Predicted	
	Confusion Matrix	True	False
Actual	True	A	B
	False	C	D

The Accuracy can be determined by

$$Accuracy = \frac{\left[\frac{A}{A+B} + \frac{D}{C+D}\right]}{2}$$

Results

a.) Unclassified Data -

After implementing the models with the selected predictors we found that even though the features are not causal they have a significant impact on the response classification, people attempting suicide and not attempting except state and year. Recursive partitioning gave most interpretable results as shown below. Here 1 is response class depicting the people who attempted suicide and 0 who did not. We can see that most important features are not only drug abuse but other features like physical fighting, being threatened at school are used for further classification.

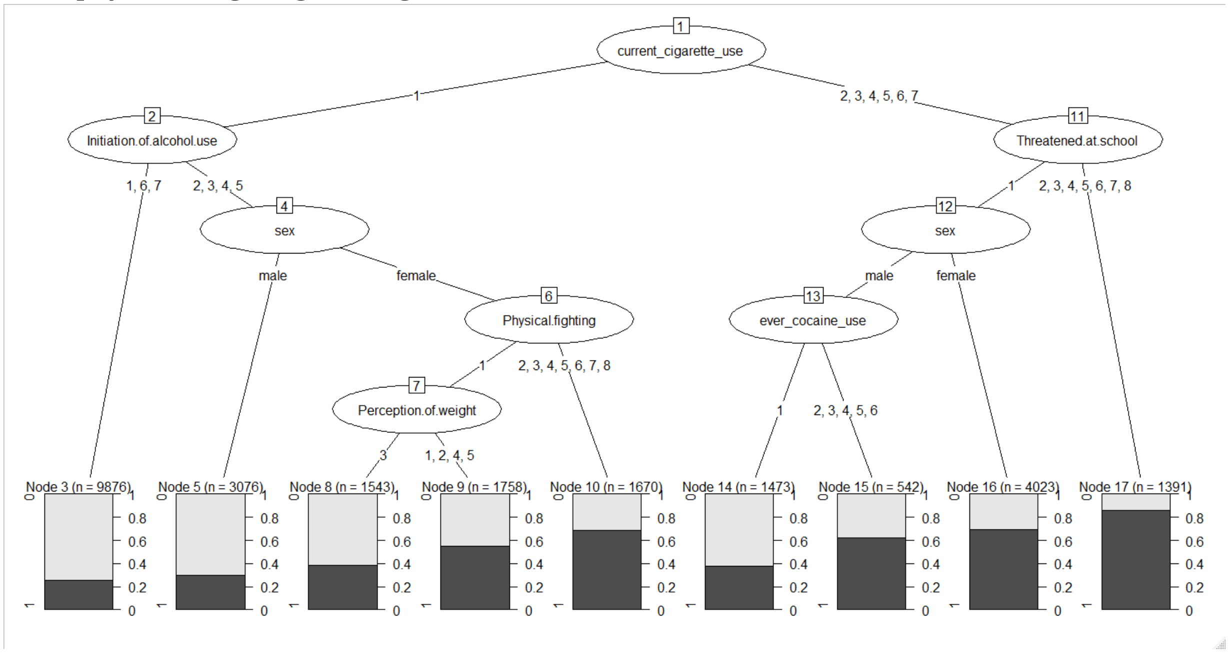

Figure 6 Classification from Recursive partitioning

From fig 7 we can see that during classification using Random Forest the individuals who did not drink alcohol but still attempted suicide were the highest. This prooves that initiation of alcohol at an early stage is not the only factor contibuting in this such behaviors among youth. From Figure 8 and Figure 7 we can see that even though initiation is an important features it is not the only factor, other factors are also contibuting in attempts.

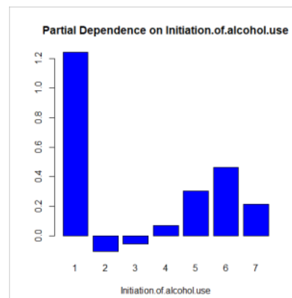

Figure 7 Partial dependence of alcohol initiation

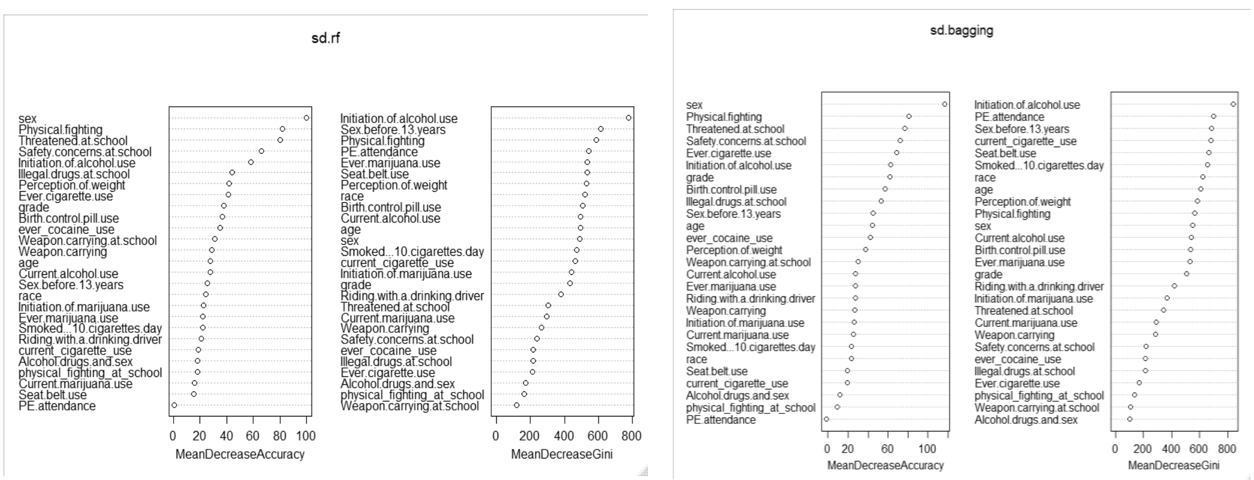

Figure 8 Variable plots from bagging and boosting

b.) Suicide Attempt Severity based on gender, male and female -

For in depth understanding the severity, it was necessary to classify the data into male and female. Same methods were used to predict the severity among the gender, male and female. In recursive partitioning, it was found that physical fighting was the most important factor further classified into different variable for different gender.

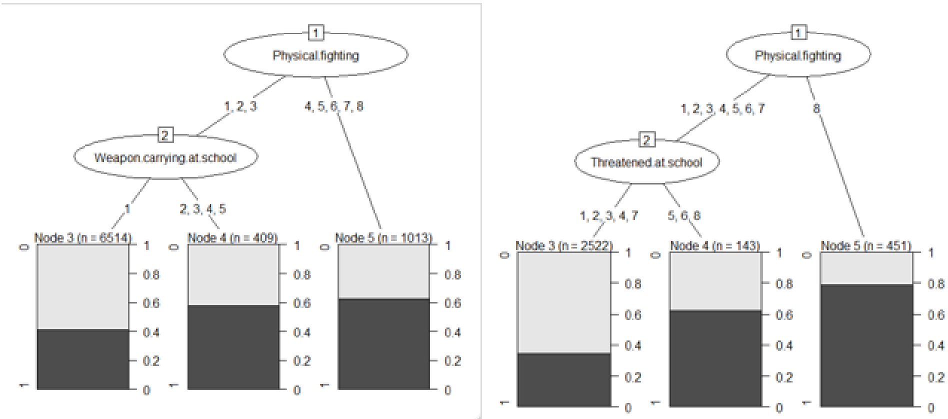

Figure 9 Male and Female Classification using Recursive Partitioning

It is classified into weapon carrying at school nad threatned at school in female and male respectively. This physical fighting means that it has happened on or off campus. According to bagging and random forest the the important variables almost remains the same i.e the top 10 predcitors. From the fig 10, it can be seen that male youths who are indulge in severe physical fighting i.e more than 10 times in past 12 months are more prone to have suicide and this makes the factor more important than anyother. For female having sexual intercourse at teenage is the important factor while partial dependence plot shows that female who had their first intercourse were in the age between 13- 17 years. These females are more likely to attempt suicide and can be said that it is one of the factors and not the only factor leading them to take such risk.

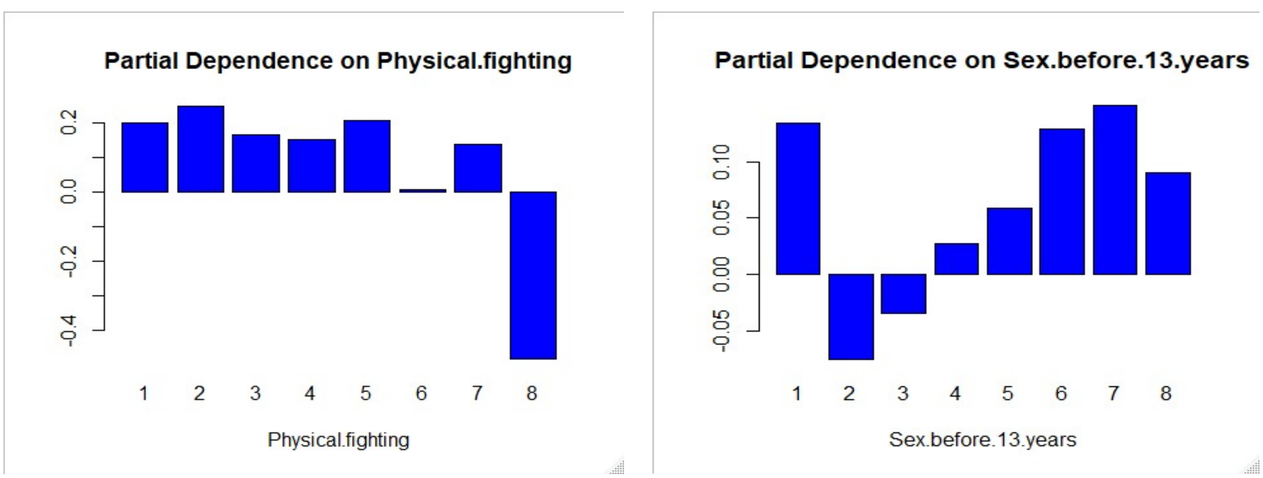

Figure 10 Partial Dependence of Important features on male and female responses

Results	Suicide Attempt Severity based on gender				Suicide Attempts	
Methods	**In-sample(Male)**	**Out-of Sample**	**In-sample(Female)**	**Out-of-Sample**	**In-sample error**	**Out of sample error**
GLM	The best fit was using classification trees due to low contrast in data				0.72	0.68
R-part	0.671	0.648	0.521	0.520	0.689	0.682
Bagging	0.994	0.667	0.518	0.506	0.987	0.7204
RF	**0.9971**	**0.659**	0.517	0.509	0.96	0.7263
SVM	0.6727	0.653	**0.592**	**0.551**	**0.97**	**0.737**

Table 3 Accuracy based on different statistical methods

The Table 3 Accuracy based on different statistical methods shows the accuracy of each algorithm derived from the confusion matrix we can see that statistical models applied to female response did not perform very well due to low variation in the data and their response due to ordinality and other factors i.e. identification of factors for severity in suicide attempts from female response was not up to par because of aforementioned reasons.

Discussions

The results obtained using supervised learning methods conclude and prove that drug and alcohol abuse are not the only factors that can be associated with suicidal tendencies. Other factors causing an individual to severe depression and anxiety disorders, feeling trapped and hopeless can be equally associated with such behaviors. Our research among youth risk factors are in support with about suicide ideations and thought. Through vigorous statistical analysis, the important features we obtained that can be linked to suicide are three main causal features like feeling hopeless and intolerable emotional pain, considering and making a plan, and various other factors like higher encounters of physical fighting and being constantly threatened at school can be linked to one's abnormal preoccupation with violence, dying or death, engaging in underage sex (irrespective of gender), one's perception of weight can be linked to feeling of agitation, shame and self-criticism. Carrying weapon, having drugs on campus, early initiations of alcohols and other narcotic substances can lead to social anxiety, panic attacks, impaired concentration and decision-making abilities.

To get even more insights on the causes and discussed factors, we need more detailed data from the survey. There is inherent problem in data collection since the survey is conducted in traditional way, lot of questions are missed and has to be removed like their dietary habits, perception of grades, sleeping patterns and lot more. The survey has to be conducted in such a way that no questions are left out while completing the survey and thus we get much details to analyze, and go to more specifics that are causing these behaviors among youths. Many important factors like in-class performance, peer pressure, family's socio-economic background and environment can be helpful to further identify the risk factors.

Conclusions

Our results show that initiation and current use of alcohol is not only factor that can be associated with suicide ideations and attempts. Factors like physical fighting, being threatened at school, lack of sexual knowledge which have them sex at underage, individual’s perception of his/her weight, not attending physical education classes are among other features that can heavily impact an individual’s psychology leading to ideations and attempts. These factors are important when classifying the response, attempting and not attempting, and while analyzing the severity of attempts we found that higher encounters physical fighting and being threatened at school are important factors. This shows that even school has an impact on youth’s mind for taking such a step. Other than family issues, one also has to identify the youth background including the school status which can help to know the mental health condition of that youth.

Future Work

Our findings were limited to gender-based classification. In future more variables such as grade points in school and high school, school attendance, extra-curricular activities etc. can also be added to the questionnaire to have an in-depth study for assessing such factors. Other than this, the study can be extended to the age group which are facing such risk including to which state. Depending upon the infrastructure, economic status, population etc. of a state can also have an effect on such sensitive topic.

References

[1] B. Walsh and D. Walsh, "Suicide in Ireland: The influence of alcohol and unemployment," Economic and Social Review, 2011.

[2] M. H. Swahn and R. M. Bossarte, "Gender, Early Alcohol Use, and Suicide Ideation and Attempts: Findings from the 2005 Youth Risk Behavior Survey," Journal of Adolescent Health, 2007.

[3] M. H. Swahn, R. M. Bossarte and E. E. Sullivent, "Age of Alcohol Use Initiation, Suicidal Behavior, and Peer and Dating Violence Victimization and Perpetration Among High-Risk, Seventh-Grade Adolescents," PEDIATRICS, 2008.

[4] L. Sher, "Alcohol consumption and suicide," QJM - Monthly Journal of the Association of Physicians, 2006.

[5] C. R. PFEFFER, J. NEWCORN, G. KAPLAN, M. S. MIZRUCHI and R. PLUTCHIK, "Suicidal Behavior in Adolescent Psychiatric Inpatients," Journal of the American Academy of Child and Adolescent Psychiatry, 1988.

[6] M. Kurokouchi, N. Miyatake, H. Kinoshita, N. Tanaka and T. Fukunaga, "Correlation between suicide and meteorological parameters," Medicina (Lithuania), 2015.

[7] D. Jacobs, E. A. Schilling, R. H. Aseltine, J. L. Glanovsky and A. James, "Adolescent Alcohol Use, Suicidal Ideation, and Suicide Attempts," Journal of Adolescent Health, 2008.

[8] Y. Im, W. O. Oh and M. Suk, "Risk Factors for Suicide Ideation Among Adolescents: Five-Year National Data Analysis," Archives of Psychiatric Nursing, 2017.

[9] J. R. Blosnich, A. J. Gordon and R. M. Bossarte, "Suicidal Ideation and Mental Distress Among Adults With Military Service History: Results From 5 US States, 2010," American Journal of Public Health, vol. 104, no. S4, pp. S595-S602, 9 2014.

[10] B. Columbia, S. A. Marion, M. O. Agbayewa and S. Wiggins, "The Effect of Season and Weather on Suicide Rates in the Elderly in British Columbia".

[11] Department of Health and Human Services. The Surgeon General's call to action to prevent suicide. Washington, DC: Department of Health and Human Services; 1999. Available online from: http://www.surgeongeneral.gov/library/calltoaction/default.htm.

[12] AFSP's latest data on suicide are taken from the Centers for Disease Control and Prevention (CDC)

Data & Statistics Fatal Injury Report for 2017. Suicide rates listed are Age-Adjusted Rates.

URL - https://afsp.org/about-suicide/suicide-statistics

[13] Centers for Disease Control and Prevention. YRBSS Online. Youth Online: Comprehensive Results. Interactive Query System Available online from:

URL: https://www.cdc.gov/healthyyouth/data/yrbs/pdf/trendsreport.pdf
URL: http://apps.nccd.cdc.gov/yrbss/ [accessed March 6, 2007].

[14] Darvishi, N., Farhadi, M., Haghtalab, T., & Poorolajal, J. (2015). Alcohol-related risk of suicidal ideation, suicide attempt, and completed suicide: a meta-analysis. PloS one, 10(5), e0126870.

[15] J. Rehm, C. Mathers, S. Popova, M. Thavorncharoensap, Y. Teerawattananon and J. Patra, Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders, 2009.

[16] Markowitz, S., Chatterji, P., & Kaestner, R. (2003). Estimating the impact of alcohol policies On youth suicides. Journal of Mental Health Policy and Economics, 6(1), 37-46.

[17] Curtin, S. C., Warner, M., & Hedegaard, H. (2016). Increase in suicide in the United States, 1999–2014.

[18] Borowsky, I. W., Ireland, M., & Resnick, M. D. (2001). Adolescent suicide attempts: risks and protectors. Pediatrics, 107(3), 485-493.

[19] Bridge, J. A., Asti, L., Horowitz, L. M., Greenhouse, J. B., Fontanella, C. A., Sheftall, A. H., ... & Campo, J. V. (2015). Suicide trends among elementary school–aged children in the United States from 1993 to 2012. *JAMA pediatrics*, *169*(7), 673-677.

[20] Centers for Disease Control and Prevention. (2007). Suicide trends among youths and young adults aged 10-24 years--United States, 1990-2004. *MMWR: Morbidity and mortality weekly report*, *56*(35), 905-908.

[21] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. *An Introduction to Statistical Learning*. Springer