COMPARATIVE ANALYSIS OF ALCOHOL CONSUMPTION PREDICTION BY USING MACHINE LEARNING ALGORITHMS

Santhiya[1] and Nancy Jasmine Goldena[2]

[1]PG Student, Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), Tirunelveli, Tamil Nadu, India.

[2]Associate Professor, Department of Computer Applications & Research Centre, Sarah Tucker College (Autonomous), Tamil Nadu, India.

Abstract:

In south Indian regions as opposed to consumption of alcohol by both men and women has increased in urban India, according to the National Family Health Survey, 2019–21. 1% of women who already are 15 or older and 19% of men in the same age range use alcohol. In India, there will be 6.5 billion litres of alcohol consumed by 2023. Alcohol-related problems such as gastrointestinal problems, intestinal troubles, high blood pressure, coronary heart disease, accidents, and several other health concerns, particularly in diabetes. In this work, the Kaggle machine learning repository's alcohol consumption dataset is analysed. This dataset includes information on the physiological activity, obesity, and other factors of diabetes patients who consume alcohol. Two machine learning techniques are used in a comparative analysis since machine learning is always successful for real-world issues. In this latest dataset, classifiers are utilised to conduct binary classification and calculate alcohol consumption for both men and women. The reliability of the identification is evaluated using a commonly used performance metric.

Keywords: Alcoholism, Classification, Health, India, Visualization, ML, Prediction.

Introduction:

Alcohol drinking is quite prevalent in the society, which results in large losses due to harmful physical health effects like cirrhosis of the liver, absenteeism, road accidents, and other mental health problems. Alcohol is a significant cause of death and disability worldwide, including in India. Alcoholism can lead to mental health issues, delirium tremens, Wernicke-Korsakoff syndrome, irregular heartbeat, a weakened immune system, liver cirrhosis, and an elevated chance of developing cancer. Prenatal drinking can cause deadly alcohol spectrum diseases. Women are typically more susceptible than men to the negative effects of alcohol, mainly because they are physically smaller, have a weaker metabolism for alcohol, and have a higher percentage of body fat. A small percentage of people experience cognitive decline and frank dementia as a result of persistent, severe alcohol use. About half of the risk of developing alcoholism is related to each of two factors: environment and

genetic. As alcohol intake can temporarily decrease dysphonia, stress and related problems, such as anxiety, are important contributors to the development of alcoholism.

Motivation of the Work:

The ability to motivate patients to visit, complete, and comply with treatment as well as successfully implement long-term changes in their drinking is crucial for the treatment of alcoholism. In terms of their willingness to change their drinking habits, both alcoholics and patients can be divided into "stages of the development process." As a result, researchers have had to take motivation's involvement in the treatment of substance misuse and recovery from it more seriously and incorporate motivational enhancement techniques into treatment plans. Heavy drinking can accumulate specific acids in the blood, which can have detrimental effects on health, especially in diabetics. Finally, drinking alcohol might worsen diabetes-related health issues such nerve damage, eye problems, and changes in fat metabolism. Machine learning algorithms were used to conduct a comparative analysis of alcohol consumption.

Dataset Description:

The dataset contains 952 rows and 17 columns. The KNN and GLM algorithms were used to compare how many diabetic patients' male and female drinks were consumed. The outline of the work is shown in **Figure 1.**
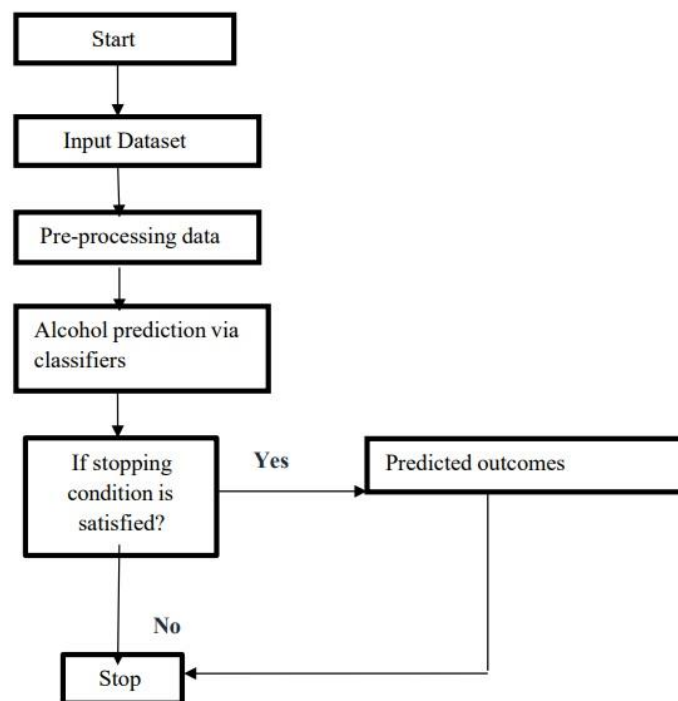

Figure 1: Outline of the work

Methodology:

Choosing features using Pearson correlation:

When creating a predictive model, the process of feature selection involves lowering the number of input variables. In some circumstances, reducing the amount of input variables might increase the model's performance while also lowering the computing cost of modelling. The test statistic that assesses the statistical association, or relationship, between two continuous variables is called Pearson's correlation coefficient. It provides details on the size of the association or correlation as well as the relationship's direction. When two or

more independent features have a strong correlation, they can be eliminated as duplicate features. When independent variables are highly linked, increases in one would result in changes in the others, which causes the model results to differ significantly. Given a little change in the data or model, the model outcomes will be unstable and fluctuate greatly. Correlations of both a positive and negative sign are factored.

Alcohol Classification:

The analysis of which data mining methodology, classifier, or system is optimal for obtaining the desired results is complicated due to the large variety of methodologies, models, and studies that have been conducted. Every system depends on a unique feature data; hence it is crucial to create a prototype that can be adjusted to each attribute dataset. They also lack effective methods for data processing and variable selection. As a result, it is highly recommended to examine classification strategies before moving on to analyse the applicant data. Additionally, classifiers are useful in accurately predicting results to determine whether alcohol use has an impact on the application. As a result, the best classifier out of the two used in this work—KNN and GLM—is selected for use in further studies.

KNN:

One of the simplest machine learning algorithms, based on the supervised learning method, is K-Nearest Neighbour. The KNN method considers that the new case/data and the existing cases are comparable, and it places the new case in the category that is most similar to the existing categories. The KNN algorithm maintains all the information that is accessible and categorises new data points based on similarity. This means that as fresh data is generated, it may be quickly and accurately categorised using the KNN method. Although the KNN technique is most frequently employed for classification problems, it can also be utilised for regression.

GLM:

Regression models are estimated using Generalised Linear Models (GLM) for outcomes with exponential distributions. These include the Poisson, binomial, and gamma distributions in addition to the Gaussian (or normal) distribution. Each has a particular function and can be used for classification or prediction based on the distribution and connection function used. The GLM extends the capabilities of linear regression by allowing the linear model to be connected to the response variable via a link function and by allowing the variance of each measurement to be a function of the projected value.

Performance Evaluation:

The performance evaluation metrics are used to calculate the performance of your trained machine learning models. This helps in finding how better machine learning model can perform on a dataset that it has never seen before. Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

Table 1 : Performance evaluation result

S.No	Classifier	Accuracy
1	GLM	82%
2	KNN	76%

Conclusion:

Detecting alcohol intake risk at an early stage is one of the major health challenges. This study makes an effort to set up a framework that predicts the risk associated with alcohol use. Two machine learning classification algorithms were used in this study, and the outcomes were contrasted using several statistical metrics. Also, the diabetes database was used with the same techniques. According to the experimental findings, the accuracy is the greatest of the group at 82.204%. For various parameters, such as precision, recall sensitivity, etc., all the models generated good results using the 2 different machine learning methods that were used. This study still has room for development, including the use of other machine learning predictors of alcohol intake.

Future Enhancement:

In future, additional diverse datasets may be used. As for now One performance evaluator is all that is employed in the analysis. In future more performance evaluation metrics may be employed. Future systems will be able to determine whether or not people who drink alcohol are harmful to others. In future most crimes in contemporary culture are committed by alcoholics who are unable to regulate both their mental and bodily emotions. In order to develop leads on criminal activity committed by alcoholics.

References:

1. World Health Organization (WHO) Hammer J.H., Parent M.C., Spiker D.A. Global Status Report on Alcohol and Health 2018. Vol. 65, Global Status Report on Alcohol. 2018.[(accessedon30March2022)].pp.74–85.Available online: https://www.who.int/publications/i/item/9789241565639

2. Eashwar V.A., Gopalakrishnan S., Umadevi R., Geetha A. Pattern of alcohol consumption and its associated morbidity among alcohol consumers in an urban area TamilNadu. J.FamMedPrim.Care. 2019;8:2029. doi: 10.4103/jfmpc.jfmpc_226_19. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

3. World Health Organization (WHO) Global Action Plan for the Prevention and Control of No communicable Disease (2013–2020) 2013. [(accessed on 30 March 2022)]. Available online: https://www.who.int/publications/i/item/9789241506236

4. Htet H., Saw Y.M., Saw T.N., Htun N.M.M., Mon K.L., Cho S.M., Thike T., Khine A.T., Kariya T., Yamamoto E., et al. Prevalence of alcohol consumption and its risk factors among university students: A cross-sectional study across six universities in Myanmar. PLoS ONE. 2020;15:e0229329. doi: 10.1371/journal.pone.0229329. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

5. Fuchs FD, Chambless LE, Whelton PK, Nieto FJ, Heiss G. Alcohol consumption and the incidence of hypertension: the Atherosclerosis Risk in Communities Study. Hypertension. 2001; 37: 1242–1250.

6. Benegal V. India: Alcohol and public health. Addiction. 2005;100:1051–6.The Hindu. Alcohol advertising in the digital age. [Last accessed on 2019 Jun 15]. Available from

https://www.thehindu.com/entertainment/music/alcohol-advertising   in-the-digital-age/article25968571.ece .

7. Wangdi K, Jamtsho T. Prevalence and correlates of current alcohol use among Bhutanese adults: A nationally representative survey data analysis. Indian J Psychol Med. 2019;41:38.

8. Ramanan VV, Singh SK. A study on alcohol use and its related health and social problems in rural Pondicherry, India. J Family Med Prim Care. 2016;5:804.

9. Manimunda SP, Sugunan AP, Thennarasu K, Pandian D, Pesala KS, Benegal V. Alcohol consumption, hazardous drinking, and alcohol dependency among the population of Andaman and Nicobar Islands, India. Indian J Public Health. 2017;61:105.

10. Bute J, Kori S, Arora VK, Sahasrabuddhe A. Prevalence and pattern of alcohol consumption among males attending primary health care setting using AUDIT in rural Indore, Madhya Pradesh, India. Int J Community Med Public Health. 2018;5:4461–9.

11. Chandra P, Krishna V, Benegal V, Ramakrishna J. High-risk sexual behaviour and sensation seeking among heavy alcohol-users. Indian J Med Res. 2003;117:88– 92.

12. Adityanjee, Mohan D, Wig NN. Alcohol-related problems in the emergency room of an Indian general hospital. Aust N Z J Psychiatry. 1989;23:274–8.

13. Alcohol & India Nationalism. A Review of A World Without Drink: Temperance in Modern India, 1880-1940 by Robert Eric Colvard. [Last accessed on 2018 Aug 25]. Available from: http://dissertationreviews.org/archives/10083 .