Prediction of the Prevalence of Substance Abuse based on the various Demographic Parameters of the Individual

Eashani Deorukhkar¹ Shreyashi Ghosh² Hemangi Ghadge³

¹²³Department of Information Technology
Ramrao Adik Institute of Technology Mumbai, India

Abstract— Substance abuse has always been a prevalent problem which has led to other problems like health issues, poverty, debt etc. This project aims to find correlations between the usage of drugs and the other variables in an individual’s life. Finding such correlations might help understand the motivations behind drug use and further, drug abuse. An aspect which will be taken into consideration is the connection between the usage or abuse of cigarettes or alcohol and the abuse of drugs. This study might help identify patterns observed in “high risk individuals”, those who are very likely to start using drugs or those who are likely to be victims of drug abuse so that the proper ameliorating measures can be undertaken before the individual starts using. We intend to make a web application which people can use to identify the probability of them being a victim of substance abuse in the future. The application can also be used by Rehabilitation centers and NGOs.

Key words: Substance Abuse, Correlations, Application

I. INTRODUCTION

Substance abuse, also known as drug abuse, is a patterned use of a drug in which the user consumes the substance in amounts or with methods which are harmful to themselves or others, and is a form of substance-related disorder. Widely differing definitions of drug abuse are used in public health, medical and criminal justice contexts. In some cases criminal or anti-social behavior occurs when the person is under the influence of a drug, and long term personality changes in individuals may occur as well.¹]

Drug abuse is no longer a current medical diagnosis in either of the most used diagnostic tools in the world, the American Psychiatric Association’s Diagnostic and Statistical Manual of Mental Disorders (DSM), and the World Health Organization’s International Statistical Classification of Diseases and ICRIS Medical organization Related Health Problems (ICD).²]

Substance abuse has been adopted by the DSM as a blanket term to include 10 separate classes of drugs, including alcohol; caffeine; cannabis; hallucinogens; inhalants; opioids; sedatives, hypnotics, and anxiolytics; stimulants; tobacco; and other substances. The ICD uses the term Harmful use to cover physical or psychological harm to the user from use. [²]

This problem is to be explored in the project.

II. LITERATURE REVIEW

Correlation is a term that refers to the strength of a relationship between two variables. A strong, or high, correlation means that two or more variables have a strong relationship with each other, while a weak or low correlation means that the variables are hardly related. Correlation analysis is the process of studying the strength of that relationship with available statistical data. [⁷]

Correlation analytics is used very often in many fields to perform studies similar to our project. Some prime examples of such studies are the influence of demographic factors on the quality of life of Parkinson’s patients [³] or the correlation between the environmental and dietary factors and the incidences of cancer and mortality in various countries. [⁴]

It has been observed that social and environmental factors are responsible for the tendency for substance abuse as was observed in Iran. [⁵]

Another trend observed was that rehabilitation programs that are tailored to a specific population have been shown to be more successful than generic programs as observed in tobacco cessation programs for people with disabilities. [⁶]

Assuming that the above trends holds true for demographic factors, we, through our project aim to find such correlations that would help identify the causes of drug abuse and hence help make better drug rehabilitation treatments.

III. SYSTEM AND METHODOLOGY

The project consists of two main parts:

- Correlation finding
- Web interface

The correlation finding is to be done using the Apriori algorithm implemented in Python 2.7. Alternately the FP-growth algorithm was also considered for the same. However the Apriori algorithm was considerably faster and more efficient while delivering results. A survey based database of drug usage habits and demographic parameters of individuals called the “National Survey on Drug Use and Health, 2013” was sourced from the Inter-university Consortium for Political and Social Research (University of Michigan, Ann Arbor).

The Apriori algorithm will generate frequent itemsets, i.e. the frequently occurring pairs or groups of items. These itemsets will further be used to find the association rules that exist in the dataset i.e. how are the items related to each other based on their occurrence. The correlations observed between the usage characteristics and demographic factors are recorded and used to formulate questions for the web interface.

The web interface, implemented using PHP and JavaScript will be in the format of a quiz with one section for demographic details and the other sections for questions for each substance to be studied. The Likert scale will be implemented to allow for a range of answers (strongly agree, agree, neutral, disagree, strongly disagree). The points for questions will be adjusted according to the correlations observed.

The results of the quiz will be displayed in the form of graphs for each substance. The demographic details will be used to display the data visualized for the particular demographic.
The Google maps API will be used to point to rehabilitation centers nearby where the individual could get help.

Another functionality implemented is IVRS calling. A call can be placed to the relevant rehabilitation center. The dataflow diagram for the correlation part is as follows:

![Data flow diagram for correlation finding](image)

The complete system design is as follows:

![System design](image)

IV. ADVANTAGES

- The correlations found can be used to develop better, cause-mitigation focused treatment plans by non-governmental organizations.
- The quiz can help individuals assess their usage levels and thus later get help if needed.
- The results can be used to create more effective government laws against illicit substance use.
- The results can be used as a base for further research in this area.

V. CHALLENGES

- The database being a survey is prone to falsification of data and thus may lead to errors in the results.
- The accuracy of the quiz results depends on the accuracy of the answers of the user.
- The correlation results have been derived from the given database and may not be relevant or accurate for every person.

VI. CONCLUSION AND FUTURE SCOPE

The process of correlation analysis is a difficult task not only from the coding point of view but also from the finding of relevant correlations point of view. This is an attempt to find correlations and the end objective is to make these correlations useful to help mitigate the problem of substance abuse.

This project could further be improved in the future by:

- More parameters- Analysis can be done by including more demographic parameters which could give us more insights. There could also be more substances included.
- Faster algorithms- The current algorithm could be optimized to give faster results for more data.

ACKNOWLEDGMENT

We thank our project guide Mrs. Anitha Senathi, project co-ordinator Ms. Reshma Gulwani and our head of department Mr. Ashish Jadhav for helping and mentoring us through this endeavour.

REFERENCES