EVALUATION OF DRUG ADDICTION CAUSES IN PUNJAB USING MACHINE LEARNING

Jaspreet Singh
Research Scholar
Department of Computer Science, Guru Nanak Dev University,
Amritsar-143001, Punjab, India

Mandeep Kaur
Project Fellow
Department of Computer Science, Guru Nanak Dev University,
Amritsar-143001, Punjab, India

Gurvinder Singh
Professor
Department of Computer Science, Guru Nanak Dev University,
Amritsar-143001, Punjab, India

,

Abstract: Nowadays, one of the most prevalent issues of every nation in the world is drug addiction. The issue of drug abuse is more common in youth of age category from 15 to 45. Punjab is no exception to this fact. This paper evaluates various causes behind the drug addiction among Punjabi youth and proposed a machine learning model to identify the key parameters for the agony of Punjab. The pre-processing of proposed model involves data extraction from web sources and cleaning of data through basic Natural Language Processing techniques. The post-processing phase contains training and testing of neural network based machine learning model for the classification of Punjabi drug abusers. The proposed method recorded highest accuracy of classification among other benchmarks available in the literature.

Keywords: drug, addiction, machine, learning

1. INTRODUCTION

Drug abuse is one of the major issues faced by developing countries like India. Heroin, opium, cannabis etc are most commonly used drugs in India. The worst outcome of drug addiction is HIV/AIDS with over 2.4 million people infected. India is third highest country in term of rate of infection in the world. The worst affected states are Mizoram, Punjab, Manipur, Assam, Uttar Pradesh. Punjab known as bread basket of India produces close to 17% of countries wheat and 12% of countries rice is suffering from major social issue called Drug abuse. It is converting from a bread basket to drug basket. In Punjab nearly 75% that is 3 out of every 4 persons are taking drug in one form or another. Heroin is the most preferred drug by addicts 53%, followed by opium, doda or bhuki 33% and rest 14%

including other drugs like alcohol, tobacco, smack, poppy, cannabis etc. Skyrocketing trend of drug abuse in Punjab has welcomed many problems like family disputes, clashes, unemployment, depression, road accidents, delinquent behavior and the worst are HIV/AIDS, increased juvenile crime, suicide etc.

2. RELATED WORK

Recently many researchers have evaluated various causes for drug abuse in distress people. They have considered the national level social impacts of drug addiction among various masses of people, the side effects of this issue, and the key type of drugs involved, the brief summary of concerned work in the literature is given in Table 1 below.

Table1: Summary of related approaches for drug abuse statistics

Author, Publication, (Year)	Description	Data Set Used	Side Effects	Drugs Considered	Most Vulnerable People	Overall Statistics
Hemraj Pal., et al., INTJCURR SCI, (2015) [1]	Evaluating trend of Drug abuse in India through NHSDA	National Household survey of Drug abuse (NHSDA)	Mental disorders, Addictive Disorders	Tobacco, alcohol, Cannabis, Opioid, Sedative Hypnotic, Hallucinogens, Stimulants,	61% of people responded uses drug in their whole life in any form	Tobacco= 56%, Alcohol= 19.6%, Cannabis= 3.8%, Opium= 0.6% Heroin= 0.2%

				Volatile substance, Heroin		
R.C.Jiloha., Delhi psychiatric society,(2009) [2]	Drug abuse in adolescents is evaluated considering varioys social and cultural aspects	UN Convention reports on Narcotic Drugs and Psychotropic substance, Control board report 2012 etc	HIV/AIDS, conflicts, traffic accidents, Juvenile crime, delinquent behavior, violence, infections, physical and mental consequences, Drug trafficking and legislation etc	Cannabis, Opium, Alcohol, Heroin, Pharmaceutical Drugs, Smoking	Prior to the age of 18, 70% adults started smoking	Some type of drug is abused by 90% street children's in the world
Laura Acion., et al., PLOS ONE, (2017) [3]	Prediction of substance use disorder treatment success using machine learning framework	National wide database of 99,013 substance use disorders patients	Criminal involvement, readmissions, unemployment, less income	Alcohol, Cocaine, Cannabis, Nonprescription opiates, methamphetamine	Results revel that males in age group(25-29) with full time employment are most addicted to drugs	Alcohol=51.3%, Cannabis=26.5%, Cocaine=8.6%, Non prescription opiates=7.9%, Methamphetamine=2.3%
RavneetKaur., IJESC, (2016) [4]	Evaluation of drug abuse in Punjab using fuzzy verdict mechanism	Survey based Dataset	Physical and mental health, unemployment, frustration, family clashes, milieu disorder	All Drugs considered	Student scoring more than 5 in Survey	Proposed Fuzzy verdict mechanism predict drugaddiction with 70% accuracy
K.Kaur., Human Biology Review, (2017) [5]	Analyzing drugs statistics in Punjab showing employed, Literate and educated male are addicted worst	Government resources, newspaper	Family dispute, divorce, debt on families, broken marriages, loss oflives, HIV/AIDS, peer pressure, depression, loneliness, unemployment, Delinquent behavior, Loss in term of human potential	Cannabis, Smack, Poppy	Literate and educated(89%), employed people(83%), Males (99%)	Percentage of household affected by drug abuse is Majha=61%, Malwa=64%, Doaba=68%

Drug abuse is a global issue. India is third highest country in the world facing this problem. Punjab state of India is worst affected from it. Hemraj Pal. et al. [1] evaluated the trend of drug abuse in India through National Household Survey and revels that 61% of people responded to the survey uses drug in their whole life in some form like tobacco, cannabis, alcohol etc. R.C.Jiloha [2]evaluated drug abuse in adolescents considering the social and cultural aspects and his analysis showed that 90% of street children consume some type of drugs and his work outperforms [5].

LuraAcion.et al. [3] uses machine learning algorithm to predict substance use disorder treatment success and his work is better than [4]. RavneetKaur [4] evaluated drug addiction in Punjab using Fuzzy verdict mechanism.

3. VALIDATION OF PROPOSED METHOD

Table2 below shows the comparison of benchmarks with problems due to drug abuse. From the table it is clear that most common outcomes of drug addictions are physical and mental health related, Juvenile crime, unemployment, conflicts as discussed by [1-5].

Table 2. Comparison of Benchmarks with side effects of Drug abuse taken as parameters, **B1**: Hemraj Pal. et al., **B2**: R.C.Jiloha, **B3**: Laura Acion. et al., **B4**: RavneetKaur, **B5**: K.Kaur

Benchmarks / Parameters	B1	B2	B3	B4	B5	Proposed
Family dispute	×	×	×	✓	✓	✓
Divorce	×	×	×	×	✓	✓
Debt on families	×	×	✓	×	✓	✓
Loss of lives	×	×	×	×	✓	✓
HIV/AIDS	×	✓	×	×	×	✓
Physical and mental health	✓	✓	×	✓	×	✓
Frustration	×	×	×	✓	×	×
Traffic accidents	×	✓	×	×	×	✓
Juvenile crime	×	✓	✓	×	✓	✓
Drug trafficking and legislation	×	✓	×	×	×	✓
Infection	×	✓	×	×	×	×
Peer Pressure	×	×	×	×	✓	✓
Depression	×	×	×	×	✓	✓
Loneliness	×	×	×	✓	×	×
Unemployment	×	×	✓	✓	✓	✓
Conflict	×	✓	×	×	✓	✓
Delinquent behavior	×	✓	×	×	✓	✓
Loss in term of human potential	×	✓	×	×	✓	✓
Milieu Disorder	×	✓	×	✓	×	✓
Addictive disorder	✓	✓	×	×	×	✓

4. WORDCLOUD OF DRUG ABUSE STAISTICS IN PUNJAB

From our proposed work it becomes clear that the most common problems in Punjab due to drug abuse are family disputes, divorce, Debt on families, loss of lives, HIV/AIDS, physical and mental health, Traffic accidents, juvenile crime, drug trafficking and legislation, peer pressure, unemployment, conflicts, delinquent behavior, loss in terms of human potential, milieu disorder, addictive disorder [6, 7] as shown in the Table2 and word cloud inn the Figure1 below.

Figure1. Wordcloud of drug addiction reports from Punjab region

Dataset used to make Word cloud is taken from newspapers and government resources. It is made in R Language. Most frequent words in dataset are displayed highlighted, bold and large font size in the center of the circle then words with somewhat less frequency are shown in the next outer layer of the concentric circle and so on.

5. PROPOSED ALGORITHM

Step1: Dataset extraction from web sources in notepad files and saves with ".txt" extension.
Step2: Preprocess the data from **Step3** to **Step7**.
Step3: Read notepad file and check its structure.
Step4: Collapse all lines of file in single line and save it with particular name.
Step5: Remove punctuation marks, numbers from file and convert to lowercase.
Step6: Remove stop words, single character, extra white spaces from file.

Step7: Split dataset string into words and convert class of dataset to character.
Step8: Take list of all positive and negative words in the world, compare with dataset and show overall statistics.
Step9: Train a neural network model using preprocessed dataset and generate wordcloud.
Step10: Testing the proposed model on test set and record the accuracy of prediction.

The proposed work is implemented with neural network algorithm. After cleaning the dataset of file saved with ".txt" extension by removing punctuation marks, stop words, white spaces it is compared with lists of all positive and negative words in the world containing 2006 and 4783 words respectively. It is found that dataset of drug abuse in Punjab contains 77 positive words and 249 negative words. It shows that more negative words are matched with database than positive words which means drugs has affected the life of Punjabi's the worst, specially family members of the person consuming drug lives a miserable life.

Table3. Validation of proposed model with benchmarks

(Acion, Kelmansky and Laan)**Technique**	**Accuracy in prediction of drug addicts behavior**
Fuzzy verdict mechanism[4]	70%
Supervised Learning[3]	Between 0.793 and 0.820
Proposed Neural Network model	84.35%

6. CONCLUSION

This paper highlights the key parameters involved in the distress of Punjabi youth affected from drug abuse. The proposed model is a neural network based classification method implemented in R language for classification of text reported for drug abuse on Internet sources. The pre-processing and labeling of training dataset is done manually by human experts, however the classification accuracy reported was highest among fuzzy verdict mechanism and supervised learning methods as shown in Table3 above. The proposed method covers maximum share of key parameters listed in Table2 above, thereby suggesting the main causes behind the distress youth of Punjab, and also confirms the weightage of key concerns as shown in wordcloud in Figure1. The future scope of this study considers the contribution of these causes exhaustively by correlating them with other social issues like Punjabi farmer suicide, depression in Punjabi youth, and rapidly growing road accidents in Punjab.

REFERENCES

[1] Pal, Hemraj, et al. "Prevalence ofDrug Abuse in India through a National Household Survey." INT J CURR SCI (2015).

[2] R.C.Jiloha. "Social and cultural Aspects of Drug Abuse in Adolescent." Delhi Psychiatric Society 12 (2009): 167-175. <http://medind.nic.in/daa/t09/i2/daat09i2p167.pdf>.

[3] Acion, Laura, et al. "Use of a machine learning framework to predict substance use diorder treatment success." PLOS ONE (2017). <https://doi.org/10.1371/journal.pone.0175383>.

[4] Kaur, Ravneet. "Prediction of Drug Addiction in Punjab with Fuzzy Expert System." IJESC (2016): 3068-3071.

[5] K.Kaur. "Drug Abuse! Problem is intense in Punjab, India." Human Biology Review 2017.

[6] http://www.business-standard.com/article/current-affairs/why-drugs-in-punjab-are-a-reality-even-if-the-censor-board-thinks-otherwise-116061000363_1.html

[7] http://alcoholrehab.com/drug-addiction/drug-addiction-in-india/