Prediction of Drug Addiction Using Supervised Learning

[1]Mohan S, [2]Mrs. Shailaja L K

[1]PG Student of MCA, Dr. Ambedkar Institute of Technology, Bangalore, India
[2]Assistant Professor, Department of MCA, Dr. Ambedkar Institute of Technology, Bangalore, India

Abstract - **Drug Addiction is one of the growing threats over the India. There are a lot of differences between a drug- addicted and non-addicted person on health condition, social life, particular life, and domestic life actions. So, steps should be taken to help drug habit with proper restorative issues. In this paper, we dig for the influential factors behind drug addiction and possible results to reduce the drug addiction rate. Utmost of the data of drug- addicted people and for non-addicted person data we've collected from different sources. All men and women aged group of 17 to 45. Our primary data set is constructed of only 200 qualitative data. We have used 5 algorithms that have been deployed including Logistic regression, Decision Tree, Random Forest, Naive Bayes, Support Vector Machine and their results are contrasted.**

Keywords: Drug, Machine Learning, Statistical Analysis.

I. INTRODUCTION

Machine learning is a growing area of computational algorithms that seek to pretend human intelligence by learning from their surroundings. Drug addiction is the worst social-profitable problem in ultramodern civilization. drug misconduct directly impacts the social and profitable perspective in India and is a veritably growing trouble also. The statistics in India for drug addiction in 2018, has more than 2.3 core drug- addicted. Drug addiction is a major problem and it has a relation to social and household behavior, principles. So, the government should comprise both the family and society which are the two most workable organizations to help drug addiction. So that we choose this topic because drug addiction is a complex bug in our society. But max people don't have enough knowledge about how people become addicted to drugs.

In Asia, India is 3rd largest country by area and the population is really high. The Eastern and Northern sides are really important because they're surrounded by large mountains. And the western side is fertile and rich plain land. The mountainous regions are basically suitable for illegal drug deals and drug business. The drug smugglers can easily hide in these hilly forests and safely transport the narcotics.

Most of them are very adolescent and teenage boys. These populations are getting a burden for our society rather than being regarded as manpower. This population needs help from us to come back to a productive life. We wanted to establish a supervised machine learning- based model that will easily identify a person is drug addicted or not addicted. Because drug addiction is actually dangerous to our society.

II. RELATED WORK

This section summarizes the related Machine Learning approaches used for predicting. We've found numerous recent machine learning approaches, many recent research works. In Decision tree classification algorithm it's used to find the structural, physical, and chemical characteristics of compounds that dispose them to cause A Drug Addiction. Decision trees are likely one of the most repeatedly popular methodologies for machine learning. Logistic Regression algorithm used to predicting the categorical dependent variable in form of classification. As we see max people are used Random Forest, Decision Tree, Logistic Regression, KNN, Naive Bayes, Neural Network, etc. To reach their final targets and some expensive programming tools, for example, WEKA, Python provides executions of the machine learning algorithms that are extensively used to model drug toxin prediction.

III. METHODOLOGY

In this section, we're going to discuss the procedures that we've followed. There have several ways that we followed.

3.1 Data Collection and Overview

In the current situation of India, drug addiction is a very common and growing hazard. Our main ideal is to organize the machine learning concepts and reduce the addiction rate. So, primarily we've collected data for 25 specific questions. Utmost of the samples for medicine- addicted people, we've collected from 'UNODC', all the samples are male and female and the aged group of 17 to 45. Our primary data set was constructed on 200 samples of 19 features. After cleaning and processing the original dataset, the final dataset was of 200 samples and 13 features. For collecting the information, we

followed a questionnaire methodology, asked for particular intervention, and looked for different aspects.

3.2 Data Pre-processing

We've collected the data of 200 people from UNODC website and other sources and in the end, we've got so important data. So, for farther processing, we wanted to clean and process our data. There were several missing data and type mismatch features. So, we removed 6 features which lead to many probabilities of missing values and put down other missing column values with their mean and frequency. Finally, our data set contains 200 samples with 13 features. We've used a label encoder to translate the data to a numeric format for further processing. We've performed a different statistical analysis to find important features among those features. The Chi- Square test has given us better accuracy in the final score. So, after finding out the score of each point we've dropped the less important column that will help the model being simpler and more important. Finally, our data set contains 200 samples with 13 features.

3.3 Classification

Classification is a supervising approach that categorizes the data into the desired number of classes. The aim of this work is to find out the factors behind drug addiction and predict a person's probability of being drug- addicted. In that manner, we can reduce the dependence rate and keep people especially teenagers off from this deadly habit. So, we've deployed 5 classifiers Decision Tree, Random Forest, Logistic Regression, Gaussian Naive Bayes, and Support Vector Machine. At last, we made a comparison of their performance based on different model evaluation criteria and find out the best suitable algorithm for this piece of the problem.

IV. EVALUATION METRICS

For executing the different machine learning models and finding the best one evaluation criteria are there. Different evaluation ways are introduced based on the confusion criteria similar as accuracy, precision, recall, and f- measure and our model evaluation is done grounded on these four evaluation criteria.

- **Accuracy:** It represents how numerous cases are rightly prognosticated with respect to total compliances. Hence, the delicacy is defined as follows
- **Precision and Recall:** Precision is the chance of affiliated cases set up among the recaptured samples, whereas recall is the chance of samples set up
- **F- measure:** Gives a way to combine perfection and recall into a single measure.

V. RESULTS AND DISCUSSIONS

We aimed at understanding the important factors behind drug addiction and also help an individual from being addicted to deadly drug. So, the first study was on finding the factors, and secondly to predict an individual that has higher chances to become drug- addicted.
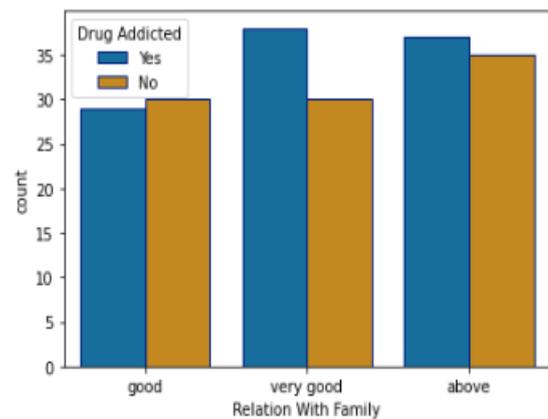

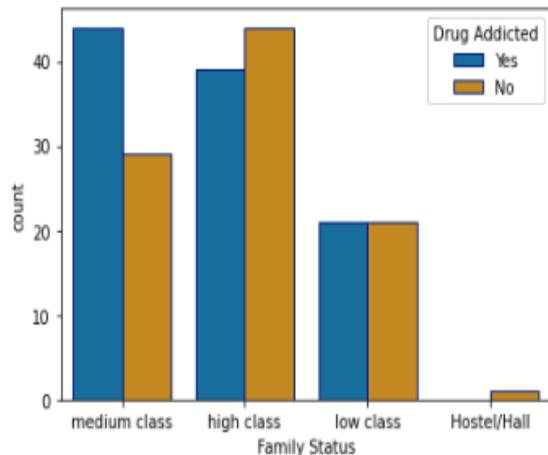**Figure 1: Drug Addiction Distribution by Family conditions**
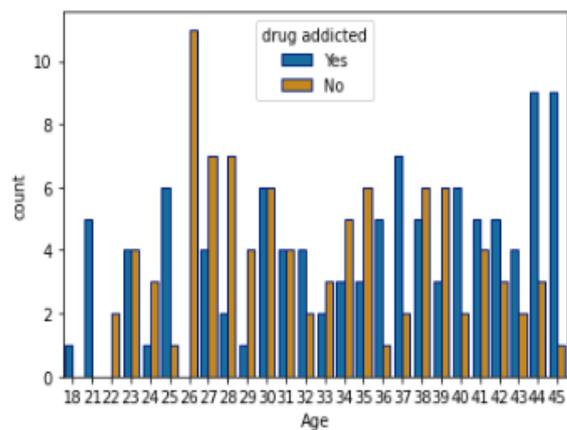

Figure 2: Drug addiction distribution by family status
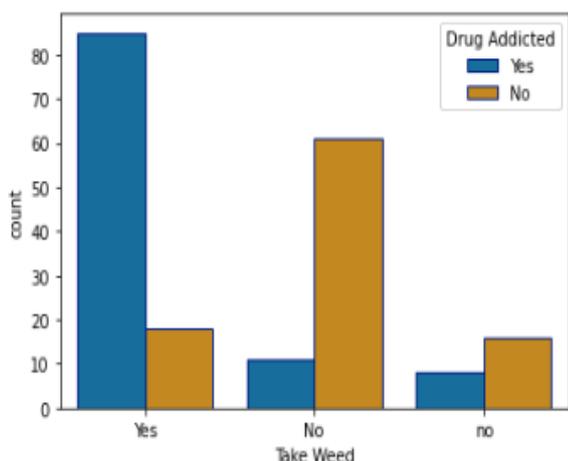

Figure 3: Drug addiction distribution by Age group
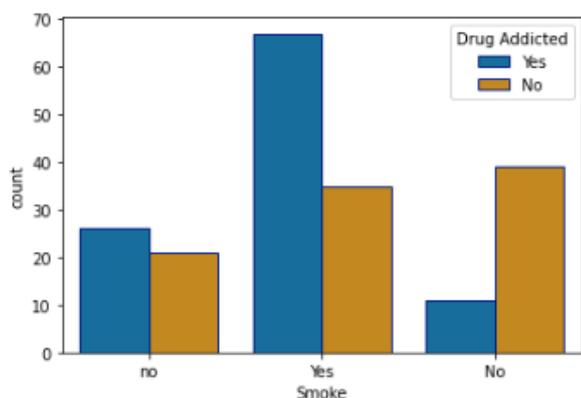
Figure 4: Distribution of whom Take Weed

Figure 5: Distribution of who Smoke

Table 1: The performances of different models

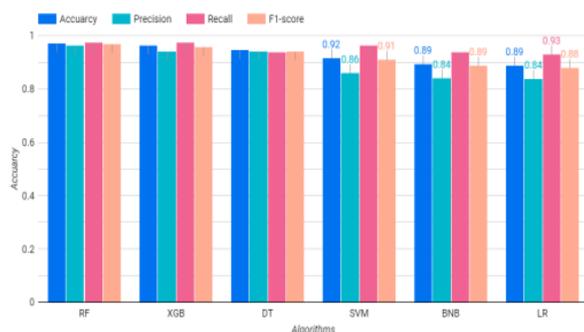| Algorithms	Accuracy	Recall	f-score	Precision
Logistic regression	0.60	0.60	0.60	0.60
Decision tree	0.45	0.44	0.42	0.40
Support vector machine	0.85	0.67	0.80	1.0
Random forest	0.80	0.91	0.83	0.77
Gaussian naïve bayes	0.80	0.67	0.75	0.86

Figure 6: Experimental Result for Different Algorithms

VI. CONCLUSION

As we have collected complete data set for males and female age groups of 17 to 45 years old. So, we expect that if future research is meditated on this topic, the area of data collection is expected to be much broader all over the country. We have some specific features. We believe we could achieve a better classification result by using some of the selected algorithms which yielded better accuracy with modified hyper-parameter tuning. Among the algorithms Support Vector Machine comes up with the highest accuracy of 85.3484, & Decision Tree Classifier delivers the accuracy of 80.2768. However, the prediction model has been developed from the 200- sample data which is not large enough to predict the class. In future research,. We will include more features and a more efficient prediction tool can be developed to get more satisfactory results.

ACKNOWLEDGEMENT

REFERENCES

[1] Hummann Felix" predicting of adverse drug reactions using decision tree modeling" Clinical Pharmacology & rectifiers88.1 (2010) 52- 59.

[2] Nath A, Kumari P, Chaube R. Prediction of human drug targets and their interactions using machine learning methods: current and future perspectives. In: Computational Drug Discovery and Design. Springer, NY, USA. 2018, 21–30.

[3] Shahriar, Arif, etal., "A Machine Learning technique to Predict of Drug Addiction" 2019 22nd International Conference on Computer and Information Technology. IEEE.

[4] Acion, Laura, etal., "Using machine learning algorithm to predict use drug treatment" PloS one12.4 (2017).

[5] Arif, Md, et al. "Drug addiction predicting Using Machine Learning", Diss. Daffodil International University, 2020.

[6] Davis Jesse., "The relationship between precision-recall and ROC curves". Conference: Proceedings of the 23rd International Conference on Machine Learning; June; 2006.

[7] Pisutaporn, Auth, Burit Chonvirachkul, and Daricha Sutivong, "Applicable factors and classification of scholar alcohol consumption" 2018.

[8] Heather, Nick., "A abstract frame for explaining drug addiction", Journal of Psychopharmacology 12.1(1998) 3- 7.

[9] Jiménez-Cordero Asunción, Maldonado Sebastián. 2018, "Automatic Feature Scaling and Selection for Support Vector Machine Classification with Functional Data", Book.

[10] Sharma H and Kumar S. "The Observation on decision-tree algorithm classification"; International journal of wisdom and exploration; 2016.

[11] Maxwell, Aaron E., Timothy A. Warner, and Fang Fang., "execution of machine- learning classification".

Citation of this Article:

Mohan S, Mrs. Shailaja L K, "Prediction of Drug Addiction Using Supervised Learning" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 6, Issue 5, pp 213-216, May 2022. Article DOI https://doi.org/10.47001/IRJIET/2022.605030

****