A Prediction Approach of Being Addicted to Drugs Using Machine Learning

Submitted by

Md. Tasnim Alam Fahim

ID: 181-35-2400

Department of Software Engineering

Daffodil International University

Supervised by

Asif Khan Shakir

Senior Lecturer

Department of Software Engineering

Daffodil International University

This Thesis report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering.

APPROVAL

This thesis titled on "**A prediction approach of being addicted to drugs using machine learning**", submitted by **Md. Tasnim Alam Fahim**, **ID: 181-35-2400** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

---------- Chairman

Dr. Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Daffodil International University

---------- Internal Examiner 1

Afsana Begum
Assistant Professor
Department of Software Engineering
Daffodil International University

---------- Internal Examiner 2

Tapushe Rabaya Toma
Senior Lecturer
Department of Software Engineering
Daffodil International University

---------- External Examiner

Prof. Dr. Md. Saiful Islam
Professor
Institute of Information and Communication Technology (IICT)
Bangladesh University of Engineering and Technology (BUET)

DECLARATION

I hereby declare that this thesis has been done by **me** under the supervision of **Asif Khan Shakir, Senior Lecturer,** Department of Software Engineering, Daffodil International University. It also declares that neither this thesis nor any part of this has been submitted elsewhere for the award of any degree.

Submitted by:

_____

Md. Tasnim Alam Fahim
ID: 181-35-2400
Batch: 25th
Department of Software Engineering
Faculty of Science & Information Technology
Daffodil International University

Certified by:

_____

Asif Khan Shakir
Senior Lecturer
Department of Software Engineering
Faculty of Science & Information Technology
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to Almighty Allah for His divine blessing makes us possible to complete the final year project/thesis/internship successfully.

We are grateful and wish our profound indebtedness to Asif Khan Shakir, Senior Lecturer, Department of SWE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Machine Learning" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Imran Mahmud, Head, Department of SWE, for his kind help to finish our project and also to other faculty members and the staff of the SWE department of Daffodil International University.

We would like to thank our entire course mate at Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURE

ABBREVIATION

LR = Logistic Regression.

SVM = Support Vector Machine.

k-NN = k-nearest neighbor.

ANN = Artificial Neural Network.

CNN = Convolutional Neural Network.

MLP = Multilayer Perception.

AUC = Area Under the Curve.

ROC = Receiver Characteristic Operator.

LASSO = Least Absolute Shrinkage and Selection Operator.

RMSE = Root Mean Square Error.

RAB = Rapid Action Battalion.

LSD = Lysergic Acid diethylamide.

DMT = Dimethyltryptamine.

ABSTRACT

Drug addiction is the incapability to refrain from consuming a legal or illegal chemical, drug, activity, or substance despite harmful consequences. It can lead to a comprehensive range of complications that harm personal relationships, professional goals, and overall health. It is one of the deadliest problems for a country like Bangladesh, where there are a large number of young people. Thus, we need to keep an eye on the young generation of our country before getting addicted to drugs. We must take efficient steps to facilitate the prevention of drug addiction. In this paper, we will predict the risk of any individual towards drug addiction using machine learning classification algorithms. First, we studied some related journals, papers and then talked to doctors, counselors, and drug-addicted people. As a result, we found some primary risk factors for addiction to drugs. Then we got a dataset from Kaggle based on the risk of drug addiction, but there was not enough data to use in the study. That's why we create a questionnaire according to each feature of the Kaggle dataset. We collected data from a couple of drug rehabilitation centers in Dhaka, Bangladesh, such as FERA Rehabilitation Center, AMI Addiction Management Institute, etc. We also collected data from a few Colleges and Universities. Our dataset includes some notable features such as age, gender, various psychological problems, lack of family ties, satisfaction in workplace or education, living with drug users, the influence of friends, and staying at a friend's house at night, etc. Our dataset contains both addicted and non-addicted samples. Our research has two outcomes: one is "Yes' means addicted, and the other is 'No' means non-addicted. After collecting the data, we processed all the data and got a processed dataset. Then we applied six machine learning algorithms to our processed dataset and compared the result of each algorithm. The algorithms we incorporated are Logistic Regression, Decision Tree, Random Forest, Naive Bayes, Support Vector Machine (SVM), and k-Nearest Neighbor (kNN). Among the algorithms, Naive Bayes came up with the highest accuracy of 90.9%, and Decision Tree delivered the least of which is 77.68%. Moreover, using a feature selection technique called chi-square, we got the most influential causes of drug addiction.

Keyword: Drug Addiction, Machine Learning, Prediction, Bangladesh.

CHAPTER 1

INTRODUCTION

1.1 Background

Drug addiction is physical and psychological incapability to refrain from consuming a chemical, drug, activity, or substance even though it is responsible for physical and mental damage. An addicted person can't control his body and mind. Because of that, when a person becomes addicted, they can't prevent the use of substances and participation in an activity. Drug addiction is one of the deadliest problems in the world. It is a vast problem for a developing country like Bangladesh. It is not recent, but it is widespread now more than ever and increasing at a terrible rate. According to the report of the Dhaka Tribune, about seven and a half million people are drug addicts in Bangladesh. Among them, 80 percent of people are teenagers and youth, of whom 50 percent are engaged in various criminal activities. In addition, around 43 percent of the unemployed population in the state are substances user [1]. The current rate of drug abuse and addiction is so high that the Government had to go hard line to stop this disaster. "Various mental problems, lack of family bonding, dissatisfaction with the workplace, stay at a friend's house at night, living with drug users, the influence of friends, habit of smoking" etc. are the most significant factors to get addicted to drugs. Its disastrous effect on society is killing our youth and our future. The effects of drug abuse are incredible. It's intense and profound. So now Bangladesh wants to get rid of drug addiction. But it is not an easy task. We need to know the actual causes of drug addiction in Bangladesh as well as the depth of the problem. We use machine learning to predict a person's hazard condition to drug addiction based on socio-economic behavior to facilitate prevention.

1.2 Motivation of the Research

Drug abuse is associated with a comprehensive range of minor and long-lasting health effects. The effects of substance abuse and dependence can be out of reach. It can damage almost every part of the human body. Drug addiction makes the addict hunger less, relentless, brutal, abnormal, anxious, intolerable, impassive, insane, incommunicative, and broken. The addicts lose their occupational and educational abilities, lose self-esteem, self-confidence, and engage in serious crime or petty criminal activity. It is very unfortunate that the solitary aim in the life of an addict becomes the collection and use of drugs. Rapid Action Battalion (RAB) caught four people in Bangladesh for the habitation of (LSD) and (DMT) drugs on June 26, 2021 [2]. For the very first time, a drug like DMT has been arrested in Bangladesh. Moreover, Oishi Rahman, a drug-addicted girl killed her parents in 2013 [3]. We need to dig up the reasons if we want to facilitate prevention. If we know the causes of addiction, we will easily prevent drug abuse as well as will be able to predict drug

addicts and non-addicts. From a Bangladeshi perspective, no significant work has been done in predicting drug addiction. That is why we are interested to work with drug addiction and machine learning techniques.

1.3 Problem Statement

Drug abuse destroys people, families, society and brings death. So drug addiction is the opponent of community and nation. Drug addict's people need our help to come back to a normal life. We wanted to establish a model by machine learning that could identify a person's vulnerability to drug addiction. Machine learning is a scheme of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from past examples, detect patterns and make decisions with minimal human intervention. There is a huge field of work in machine learning. Therefore, we thought we should apply machine learning to predict drug addiction.

1.4 Research Questions

- Which factors are most responsible for drug addiction?
- Is there any other way to identify drug abusers?
- Why should we use machine learning methods to identify the causes of drug addiction?
- Have we been able to collect the data we need for research?
- How much data have we collected and where did it come from?
- Does our collected data useful for machine learning?
- Which machine learning algorithm should we use to identify drug addicts and non-addicts?

1.5 Research Objectives

- By using machine learning techniques we will identify and validate the primary risk factors of drug addiction.
- There is little work has done to identify drug abusers. Therefore, there is not much effective way to identify drug addicts. We hope our research will help society to identify drug addicts by their socio-economic behavior to facilitate prevention.
- Nowadays, Machine learning is more effective for data analysis, data visualization, and model building. In machine learning, systems can learn from past examples and make decisions with minimum human interference.
- We have collected numerous data to do the research properly.
- We collected a total of 602 data from rehabilitation centers, schools, colleges, universities, and the Kaggle website.

- We have collected our data in a way that is useful and compatible with machine learning techniques.
- After collecting the data, we realized that the popular techniques of machine learning are suitable for our research.

1.6 Research Scope

Our research will be beneficial for our society, especially for the youth. Therefore, guardians can take care of their teenager's boys and girls. It will reduce the frequency of drug abuse. Since doctors and psychologists medicate drug addicts, our study will help them a lot. Doctors and counselors can make people aware of the risks of drug addiction. When we move to a new place, we haven't any idea about the environment there. Therefore, there is a possibility of accidentally mixing with drug addicts. We have compiled our dataset and questionnaires based on a person's daily life and socioeconomic status. Through this research, we'll be able to detect a person in advance through his behavior whether he is addicted to drugs or not. Also, we figured out the most significant factors of drug addiction that can tell us about a potential drug abuser. Thus, it will be helpful for those individuals who have a little spark of light which keeps them away from this curse and inspires them to get around.

1.7 Thesis Outline

The overall report of this research paper is given below:

- Chapter-1 explains the background and motivation behind this research which inspired the authors to study this exceptional problem. The purpose and scope of the study have also been discussed here.

- Chapter-2 discusses some of the previous work similar to this study and mentions comparative analysis and summary.

- Chapter-3 contains a detailed description of the data collection procedure, preprocessing of data, the workflow of this research, and statistical analysis.

- Chapter-4 represents a detailed explanation of the experimental result and analysis which includes the comparative study of algorithms based on performance.

- Chapter-5 concludes the paper with limitations and future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Background

We have had to read many previous papers to solve the problem of our research work. We have tried to read research papers on drug addiction that are somewhat similar to our research. In this section, we are going to discuss the work and outcome of all those papers. At the same time, we will highlight the comparisons of those papers.

2.2 Related Works

Ahnaf Atef Choudhury et al. [4] has worked on depression to check whether depression can be effectively predicted, with the help of concerned features. The purpose of their research is to identify depression in its early stages. They collected the data by a survey which was created after consultation with psychologists, counselors, and experienced professionals. They did this research among the undergraduates of Bangladesh. They collected a total of 577 data instances for the target variable as either 'Yes' or 'No' meaning, whether an individual might be depressed or not. Based on the final label and 20 features, they applied different algorithms and compared them based on accuracy and other metrics. They used three algorithms to train and test the dataset. Random Forest was the best algorithm with 75% accuracy and better precision, recall on their research. Random Forest also has lower false negatives. And the other two algorithms K-Nearest-Neighbor, and SVM had 67% and 73% accuracy, respectively.

Alireza Amirabadizadeh et al. [5] has proposed a prediction model to identify risk factors for drug abuse. It was a potential cross-sectional study conducted at South Khorasan Province in Iran. There were a total of 678 qualified subjects for building decision tree and multiple logistic regression model of whom 70% of data were for training purposes, and 30% were only for testing purposes. They collected data based on the demographic characteristics and history of drug use. The key independent variables of their dataset included first substance experience, age at first drug use, age, place of residence, history of cigarette use, and occupational and marital status. For the decision tree model, they found the sensitivity, specificity, and accuracy at 66%, 75%, and 69%, respectively, while the Logistic Regression model was somewhat less effective at 60%, 73%, and 65%. They also analyzed the predictive performance by receiver operating characteristic (ROC) curve for different models. They found Decision Tree is the most effective model for their research.

Jeeyae Choi et al. [6] has proposed nicotine addiction prediction models using machine learning algorithms. Their study was to dig up predictive variables such as social, emotional, and environmental determinants that cause nicotine addiction among youth who use e-cigarette or hookahs. They collected data from 6,511 participants through a survey based on their use of e-cigarettes or hookah. They established their model by Random Forest with ReliefF and Least Absolute Shrinkage and Selection Operator (LASSO). In their research, most essential predictor variables were identified by ReliefF, and the optimal number of predictors was evaluated by Davies– Bouldin clustering for Random Forest. They evaluated the performance of prediction models by Root Mean Square Error (RMSE) and Confusion Matrix. In their research, noble predictors were found such as 'witnessed e-cigarette use in their household' and 'perception of their tobacco use.' Also found, 73% accuracy for Random Forest and LASSO had 63% accuracy, and Root Mean Square Error (RMSE) (SD) is 0.7436 (±0.0401) for Random forest and 0.7509 (±0.0287) for LASSO. If other machine learning classifiers and feature selection methods were used, there may have been different predictor variables showing different appeasement performances.

YoungJin Choi et al. [7] has developed drug intoxication fatality prediction models, and compared machine learning models with traditional logistic regression. They extracted 8,937 samples from the Korea Centers for Disease Control and Prevention to construct their dataset. They trained, validated, and tested each model by their dataset and compared the performance of each model. They evaluated their model on three measures such as, brier score, calibration slope, and calibration-in-the-large. In the paper, the factors that affect mortality from drug intoxication were analyzed by the chi-square technique. The Chi-Square test was performed on such characteristics, age, toxic drugs, severity, risk factors, mood and intent, and the results showed that mortality rates varied significantly among most of these variables. The multilayer perceptron model (MLP) had the best area under the curve (AUC). And the Brier score was the lowest in the phases of training and validation for MLP. Whereas the logistic regression model delivered the highest AUC score (0.827) and worst Brier score (0.0307) at the testing phase. Also, MLP achieved the second-highest AUC (0.816) and second-lowest Brier score (0.003258) in the testing phase, demonstrating better performance than the decision tree model.

Mohaamad H Afzali et al. [8] has worked on the importance of prediction modeling of adolescent alcohol use. They predict different levels of alcohol use in mid-adolescence. They worked on Canadian and Australian samples separately. They used seven machine learning algorithms and compared the performance between them. To display the predictive power of the best-performing machine learning algorithm they also did a cross-cultural study scheme in the training, validation, and test process. The performance measurement was done based on those measures: F1 prediction score, accuracy, precision, recall, negative predictive value, and area under the curve. In their research, the elastic-net machine-learning algorithm showed the best predictive performance in both included Canadian (AUC=0.869±0.066) and Australian (AUC=0.855±0.072) samples.

Yupu Zhang et al. [9] has proposed a model to predict daily smoking time based on decision tree machine learning algorithm. They used the data which was collected by the Chinese center for disease control and prevention from the information of smokers. Due to the small number of features, they proposed a feature information extraction module to solve the problem. In their research, they tested different machine learning algorithms, and in the end, came to the conclusion that the prediction model performance based on the accuracy of XG Boost with the feature extraction model is 84.11% which is better than other models. And its training was much faster than other machine learning algorithms they used.

Divya Kumari et al. [10] has proposed an approach of prediction of alcohol user based on artificial neural network (ANN)**.** They designed two ANN modules. To predict a person is an alcohol user or not, they designed ANN-D. And they also designed ANN-C to predict when alcohol is used. They considered many features for their research like age, gender, country, ethnicity, education, neuroticism, openness to experience, extraversion, agreeableness, conscientiousness, impulsive, sensation seeing, etc. By the input features ANN-D module predicts the alcohol user and ANN-C module predicts the use of alcohol on the basis of time. The findings of their research are ANN-D had 98.7% and ANN-C had 49.1% accuracy.

Sivan Kinreich et al. [11] has worked to identify people at risk for developing alcoholism disorder (AUD). They collected 656 participants' data from the Collaborative Study of the Genetics of Alcoholism (COGA). The age range of the data was about 12–30 years. In their study, they included data from six different collection sites. The data included offspring and non-offspring of European American and African American ancestry. They did Feature selection and model estimation and validation separately for each group. They used the regularization method to control overfitting and used the least absolute shrinkage and selection operator (LASSO) for feature selection. In their research, genetic data and EEG data had better accuracy.

Dhiraj Dahiwade et al. [12] has proposed a general disease prediction system based on the syndromes of the patient. In their dataset, they used disease evidence of the patient. Their dataset was downloaded from UCI machine-learning website. They used K-Nearest Neighbor (KNN) and Convolutional neural network (CNN) algorithm for this prediction. They applied two algorithms and on the basis of accuracy and time, they found 84.5% accuracy in CNN, which was better than kNN algorithm.

Ahmed M. Alaal et al. [13] has developed a model for predicting disease risk of cardiovascular on Biobank participants by machine learning. Their model can predict CVD risk. They used Auto Prognosis algorithmic tool to create their model. Based on 473 variables their model can predict CVD risk. Their dataset consists of some features like walking, health rating, diabetes, breathing, age, gender, systolic blood pressure, smoking status, hypertension, diabetes, and BMI, etc. Their proposed model had AUC-ROC: 0.774 from Auto Prognosis.

2.3 Comparative Analysis & Summary

The comparison of some related research work is given below in Table 2.1

TABLE 2.1: SUMMARY OF RELATED RESEARCH WORK

SL	Author name	Methodology	Description	Outcome
1.	Ahnaf Atef Choudhury, Md Rezwan Hassan Khan, Nabuat Zaman Nahim, Sadid Rafsun Tulon, Samiul Islam, Amitabha Chakrabarty	Random Forest, K-Nearest Neighbor, and Support Vector Machine.	Predicting depression among Bangladeshi undergraduates using machine learning.	Random Forest had the best accuracy of 75%.
2.	Alireza Amirabadizadeh, Hossein Nezami, Michael G. Vaughn, Samaneh Nakhaee & Omid Mehrpour	Decision tree and Logistic regression.	Identifying the risk factors for drug use in an Iranian treatment sample: a prognosis method using Decision Tree.	For the decision tree model, the sensitivity, specificity and accuracy were 66%, 75% and 69%, respectively.
3.	Jeeyae Choi, Hee-Tae Jung, Anastasiya Ferrell, Seoyoon Woo, Linda Haddad	Random Forest with ReliefF and Least Absolute Shrinkage and Selection Operator (LASSO).	Nicotine addiction prediction model based on young e-cigarette and water pipe users by using machine learning algorithms.	Random Forest had 73% accuracy and LASSO had 63% accuracy and Root Mean Square Error is 0.7436 (±0.0401) for Random forest and 0.7509 (±0.0287) for LASSO.
4.	YoungJin Choi, YooKyung Boo	Logistic Regression, Decision Tree, Multilayer Perception.	Comparison of Logistic Regression models with alternative machine learning methods for predicting drug addiction death risk.	The MLP had the highest area under the curve (AUC), and lowest Brier score in training, while the logistic regression showed the highest AUC (0.827) and

				lowest Brier score (0.0307) in the testing phase.
5.	Mohaamad H Afzali, Matthew Sunderland, Sherry Stewart, Benoit Masse, Jean Seguin, Nicola Newton, Maree Teesson, Patricia Conrod	Logistic Regression, SVM, Random Forest, Neural Network, Lasso, Ridge Regression, and Elastic-net.	A machine learning approach to predict adolescent alcohol use: a cross-study, cross-cultural validity.	The elastic-net machine-learning algorithm showed the best predictive performance in both samples.
6.	Yupu Zhang , Jinhai Liu , Zhihang Zhang , Junnan Huang	XG Boost, Decision tree.	Daily smoking behavior prediction based on Decision Tree algorithm.	84.11% accuracy with depth 5 in XG Boost decision tree.
7.	Divya Kumari, Sumran Kilam, Priyanka Nath, Aleena Swetapadma	ANN-D, ANN-C.	Predicting alcohol abuse in individuals using Artificial Neural Networks.	ANN-D had 98.7% and ANN-C had 49.1% accuracy.
8.	S. Kinreich ,Jacquelyn L. Meyers, Adi MaronKatz, Chella Kamarajan, Ashwini K. Pandey, David B. Chorlian , J. Zhang, G. Pandey, Stacey Subbie-Saenz de Viteri, Dan Pitti, Andrey P. Anokhin, Lance Bauer, Victor Hesselbrock, Marc A. Schuckit, Howard J. Edenberg, B. Porjesz	Regularization method, LASSO.	Predicting the risk of alcoholism disorder using machine-learning technology.	Genetic data and EEG data had better accuracy.
9.	Dhiraj Dahiwade,Prof. Gajanan Patle, Prof. Ektaa Meshram	K-nearest neighbors (kNN), CNN.	Designing disease prognosis models using machine learning methods.	84.5% accuracy in CNN.
10.	Ahmed M. Alaal, Thomas Bolton, Emanuele Di Angelantonio , James H. F. Rudd, Mihaela van der Schaar	Auto Prognosis.	Risk prediction of cardiovascular disease using automated machine learning.	Auto Prognosis had 0.774 AUC-ROC and increase accuracy.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The aim of this research is to establish a model to find out whether a person is addicted to drugs or not, to the release of becoming drug-addicted. This model is built based on a combination of people's daily life and socio-economic behavior. We have applied several algorithms to create this model. The algorithms we have used are Logistic Regression, Decision Tree, Random Forest, SVM, Naive Bayes, and K-Nearest-Neighbor. We used those algorithms for classification purposes. To build this model, we used twenty-three primary risk factors that are highly correlated with drug addiction. We analyzed the data and illustrated each feature in distinction to the results of this study. We did data preprocessing before implementation. We calculated the accuracy, precision, recall, and F1 score for each algorithm and selected the appropriate one for our model. We found Gaussian Naive Bayes had the best accuracy and was appropriate for our proposed model.

3.2 Data Collection and Dataset

First, we identified the root causes of drug addiction by reading various newspapers, journals, magazines, and articles. After that, when we started data collection we didn't understand where we could get information about drug addicts. We first wanted to collect data from people in our neighborhood, because there are many people around us who secretly use drugs. But many did not agree to give their data. Next, we go to some Drug Rehabilitation Centers to collect our data. We have collected the most data on drug addicts from Drug Rehabilitation Centers. "Phera Addiction Management & Rehabilitation Center" has helped us with all the information about drug addicts we need. In addition, we collected data from "Ami Addiction Management Institute." The Chairman of "Ami Addiction Management Institute" was very Cooperative to give us information about drug addicts. Moreover, we also collected data from different universities and colleges through a Google Form questionnaire. We get most of our non-addicted data through Google Forms. We also included some data from the Kaggle community [14]. The questionnaire of our Google Form is made up based on the data we found in Kaggle. We've talked to doctors and counselors about the causes of drug addiction to make sure our created questionnaire is accurate. Based on these questionnaires we collected data from Sylhet Agricultural University, International University of Business Agriculture and Technology, Daffodil International University, and Sylhet MC College. Finally, we were able to collect a total of 602 data, including 362 drug addicts and 240 non-addicts. We have arranged our dataset with all the necessary and changeable information.

We collected our research data based on the following parameters/ features/ factors:

- Age.
- Gender.
- Education.
- Live with.
- Spend Most Time.
- Family Relationship.
- Financial Condition of family.
- Addicted Person in Family.
- No. of Friends.
- Staying at a Friend's House at Night.
- Living with Drug User.
- Satisfaction in Workplace/ Education.
- Mental/ Emotional Problem.
- Failure in Life.
- Suicidal Thoughts.
- Smoking.
- Friend's Influence
- Ever Taken Drug.
- Enjoyable with How Much Drug.
- Addicted or Not Addicted.
- Withdrawal Symptoms
- If Chance Given to Taste Drugs.
- Easy to control the use of the drug.
- Motive/ Opinion about Drug Addiction.

We talk to some physicians, visit a few websites, and read some articles to find out about these risk factors [15], [16], [17], [18], [19].

3.3 Proposed Methodology

Figure 3.1 shows the steps of our proposed methodology.

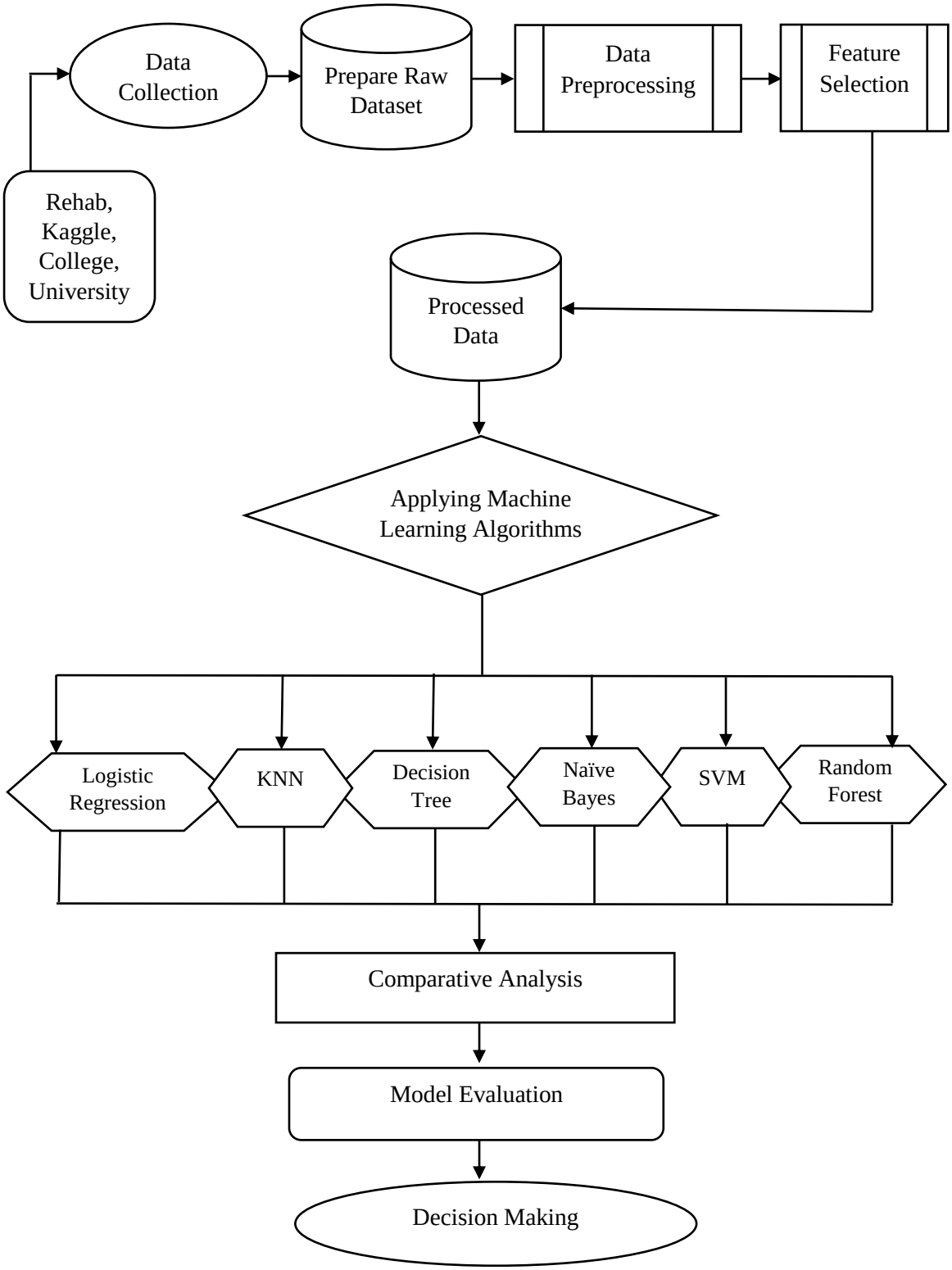

Figure 3.1: Steps of our Proposed Methodology

3.4 Data Preprocessing

After collecting the data, we get most of the data as categorical. Then we decided to do data preprocessing. Through data preprocessing, we have converted our categorical data into numerical formats so that they are suitable for working on machine learning algorithms. This is because data preprocessing has the ability to convert data which makes it suitable for application in Machine Learning algorithms.

Our Data Preprocessing scheme is shown below in Figure 3.2.

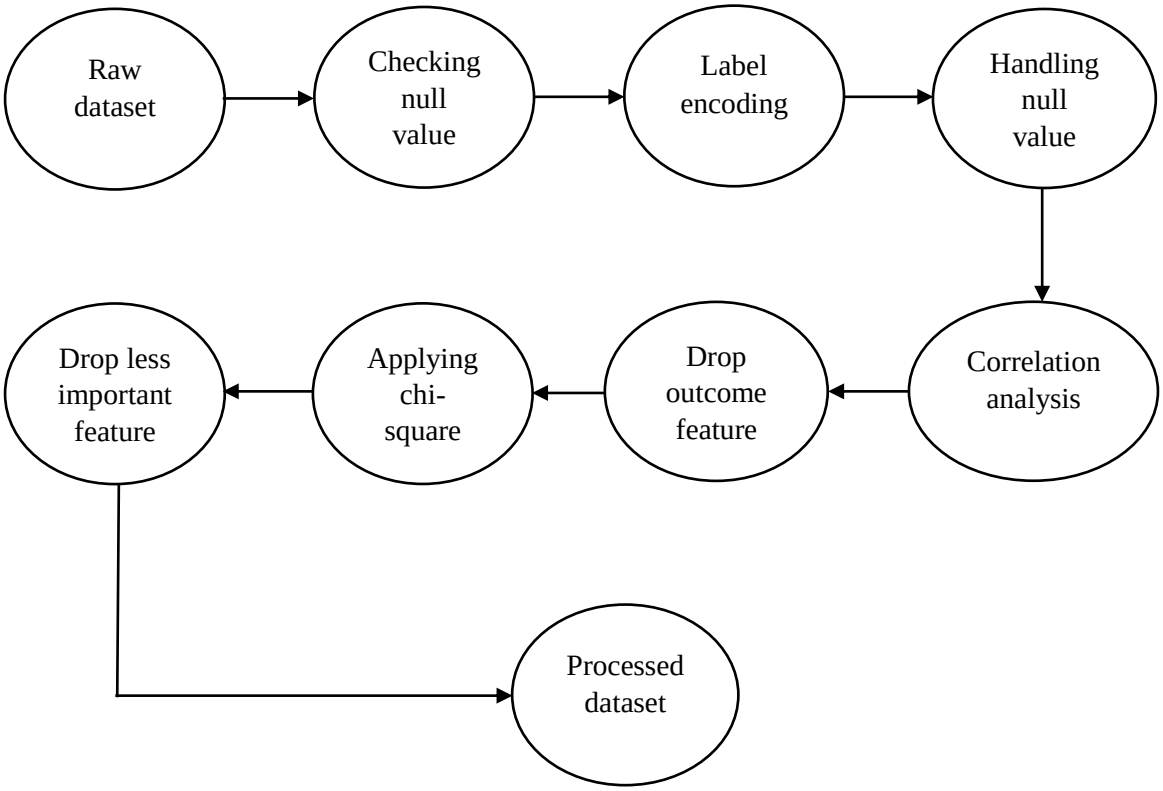

Figure 3.2: Steps of Data Preprocessing

We do the work of data cleaning first. We created many versions of the dataset using certain rules to clean the data. We first collected a total of 678 data through Google Forms and writing on paper from Drug Rehabilitation Center and Colleges, Universities, and Kaggle community. But we wanted to keep the ratio of addicted and non-addicted to 60% and 40% for our research. That's why our data number has come down to 602. To complete our Google Form survey every question had to be answered to submit that information. Therefore, when we examined the missing values, there were no missing values in our dataset. So, we had 602 data instances and this dataset was used for the prediction model. Since most of the data we collected was categorical, we converted our categorical data to numeric format with label encoding. Then we did some correlation analysis. We decompose the correlation matrix as a process of data integration. The correlation matrix shows us the ratio of each data connected to others. Then we drop the outcome feature of our dataset which was "Addicted or not addicted." After dropping the outcome feature, we

applied a feature selection technique called chi-2 to know the most important and less important risk factors of drug addiction. Therefore, we found which factors are less important in our research. Then we drop the less important feature of drug addiction. As a result, the features of our dataset have dropped to twenty. Finally, we get the processed dataset with 20 features and 602 samples on which we applied our six prediction algorithms. We have done all the steps of preprocessing with the help of "Anaconda Navigator" and "Jupyter Notebook."

There were multiple classifications based on the final score of the two addiction evaluation scales. However, as we only wanted to find out whether a person is addicted or not we removed multiple classifications. We considered the frequency of drug usage is regularly and Once/twice in a week as Yes, meaning addicted and all the other classifications as No, meaning non-addicted.

3.5 Statistical Analysis

We used 602 data in our study. Of these, 362 are drug addicts and 240 are non-addicts. That means 60% drug-addicted and 40% non-addicted. We have included data from all ages and professionals in our research.

By using a feature selection technique called chi-square, we found the most influential causes of drug addiction. From the result of chi-square, we identified 20 features as some significant risk of addiction and we only keep them on our dataset. Therefore, in this section, we will visualize some of them.

The following figure 3.3 shows how many drug addicts and non-addicts have been included in our study.

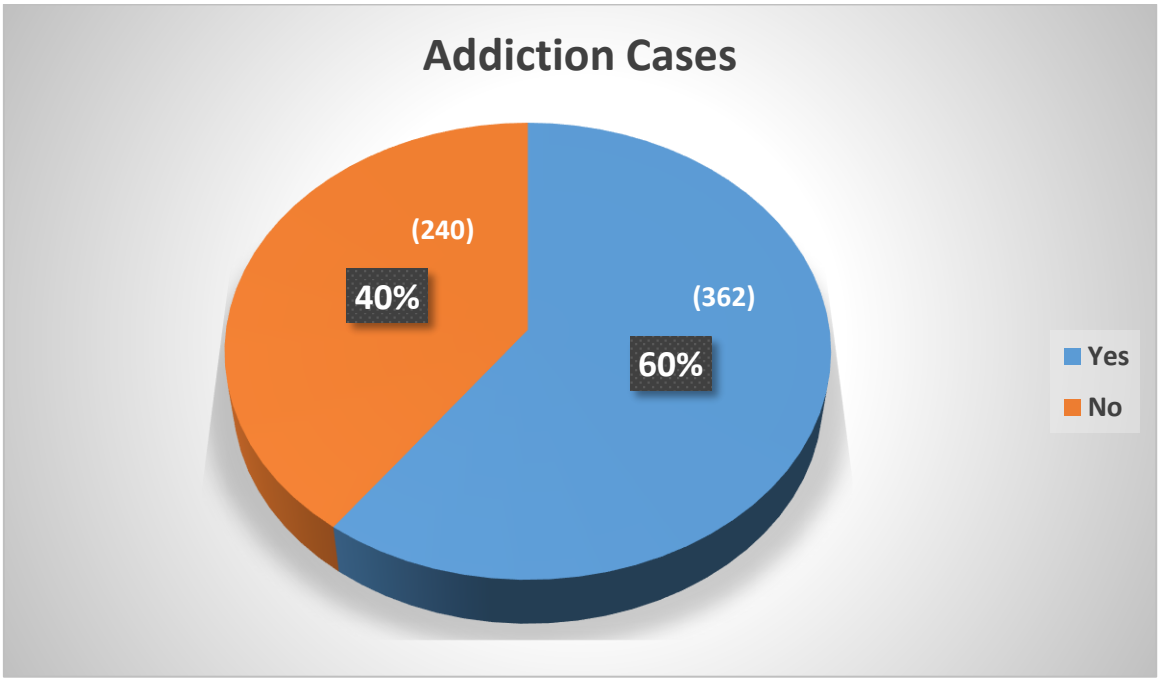

Figure 3.3: Addicted and Non-addicted Cases

The following Figure 3.4 shows the age range of the data we have collected for our study. There we see that the age of most addicts is 23 to 35 years and then 15 to 22 years.

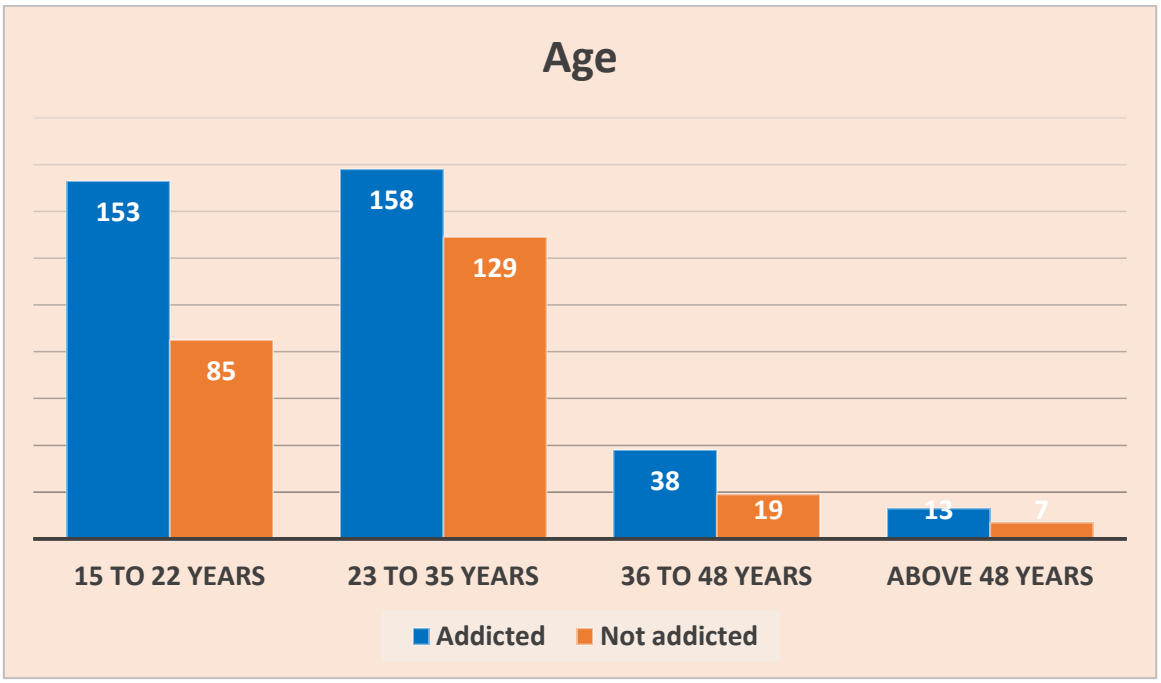

Figure 3.4: Age and Addiction Case

From the following figure 3.5, we can see, men are more likely to be addicted than women. And here are the numbers of all the men and women included in our study.

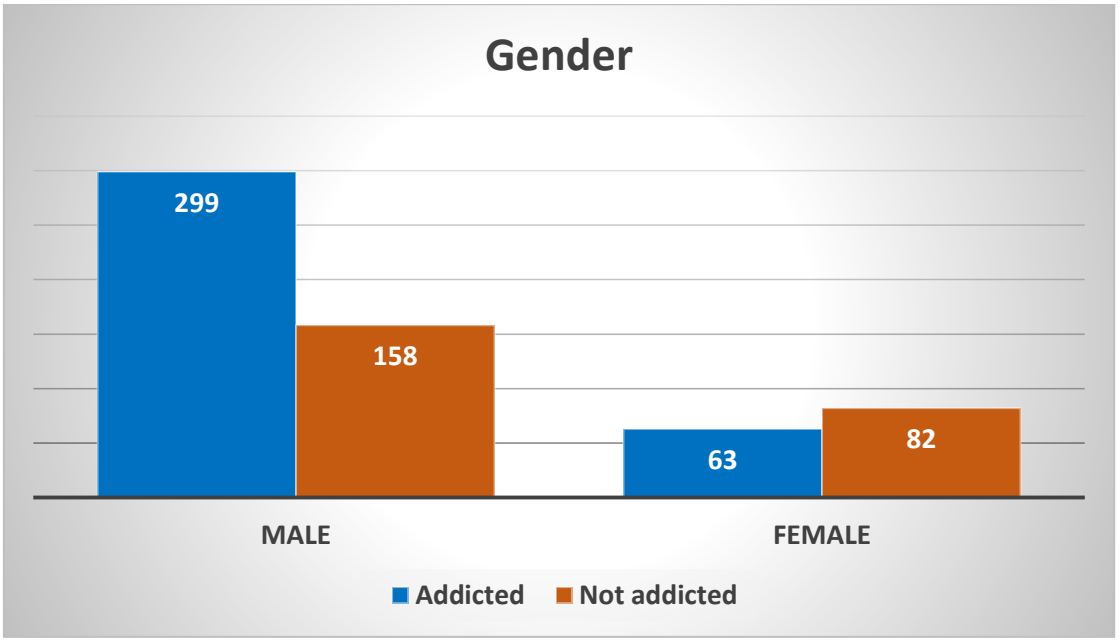

Figure 3.5: Gender and Addiction Case

As Figure 3.6 shows, family ties for drug addicts is too weak. Most addicts have a communication gap and average family bonding with their families.

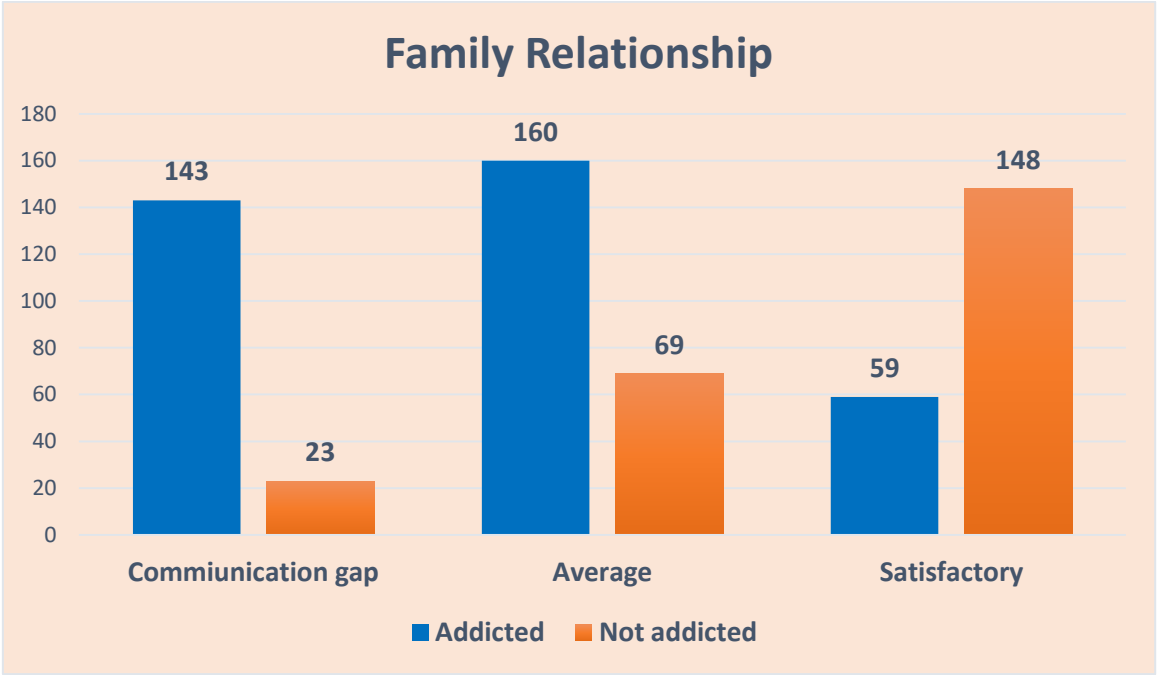

Figure 3.6: Family relationship vs Addiction

Figure 3.7 shows that those who live with drug users are more likely to become addicted.

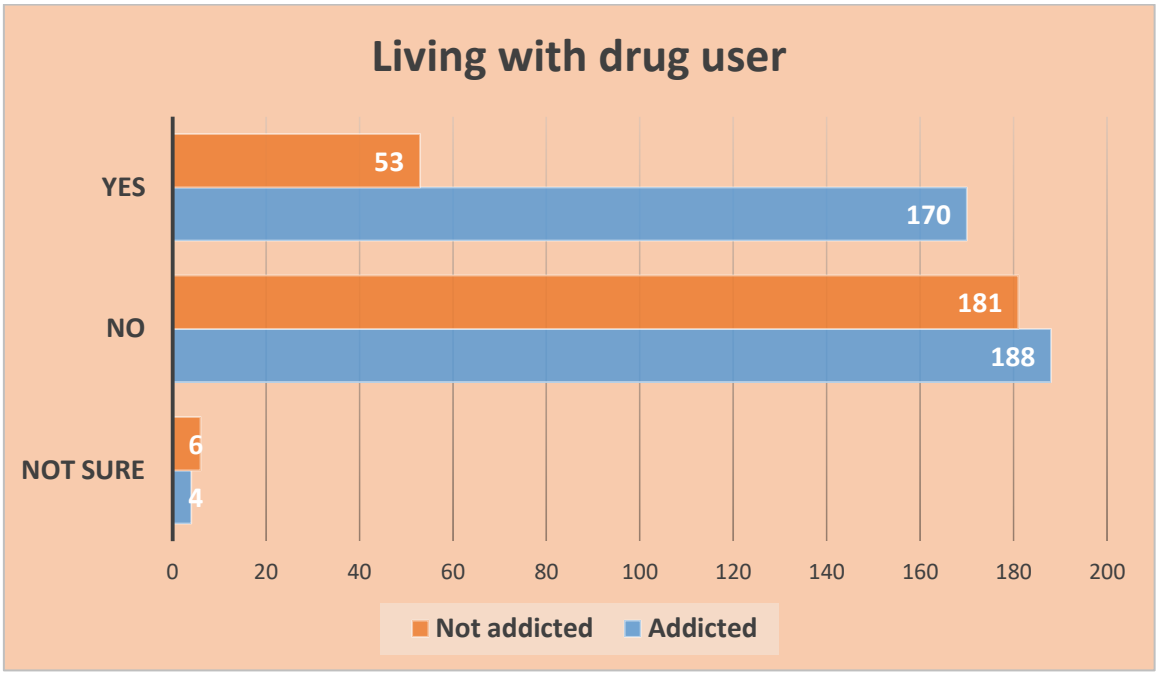

Figure 3.7: Living with drug user vs Addiction

In Figure 3.8, we see that the influence of friends is one of the most crucial reasons for drug addiction. According to our research, the rate of drug addiction is higher due to the influence of friends. On the other hand, those who are less influenced by friends are not addicted.

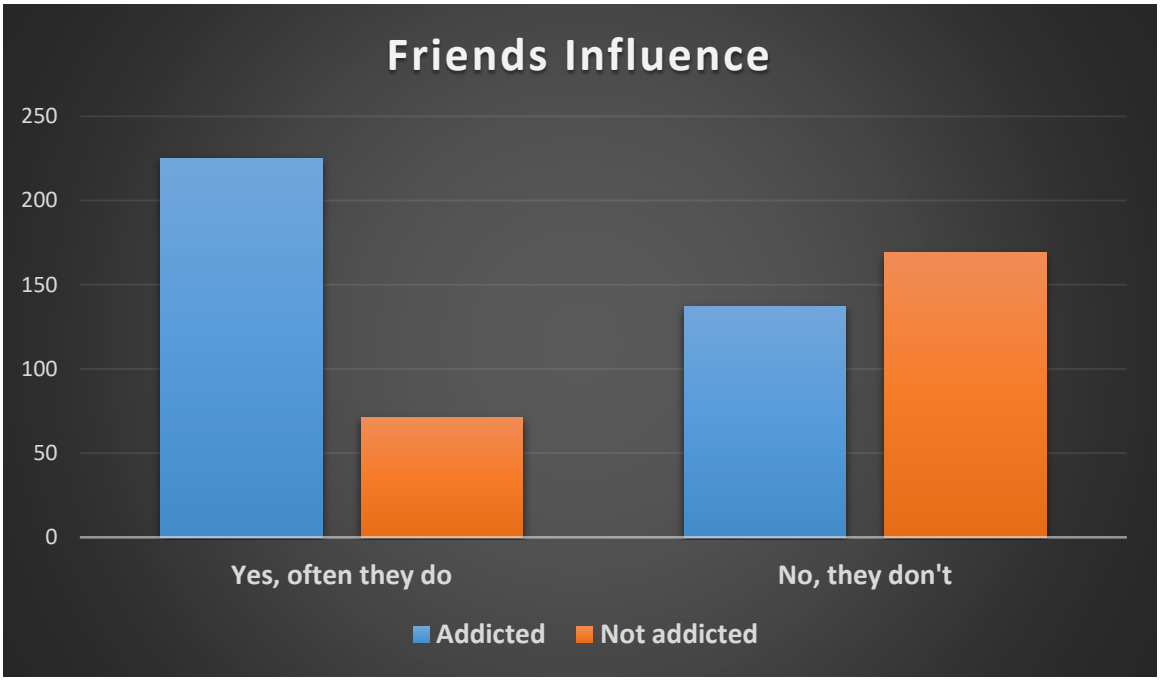

Figure 3.8: Friends influence vs Addiction

From Figure 3.9, we see that almost everyone has some mental/emotional problem. But the drug addicts have a higher level of depression, inferiority, and guilt.

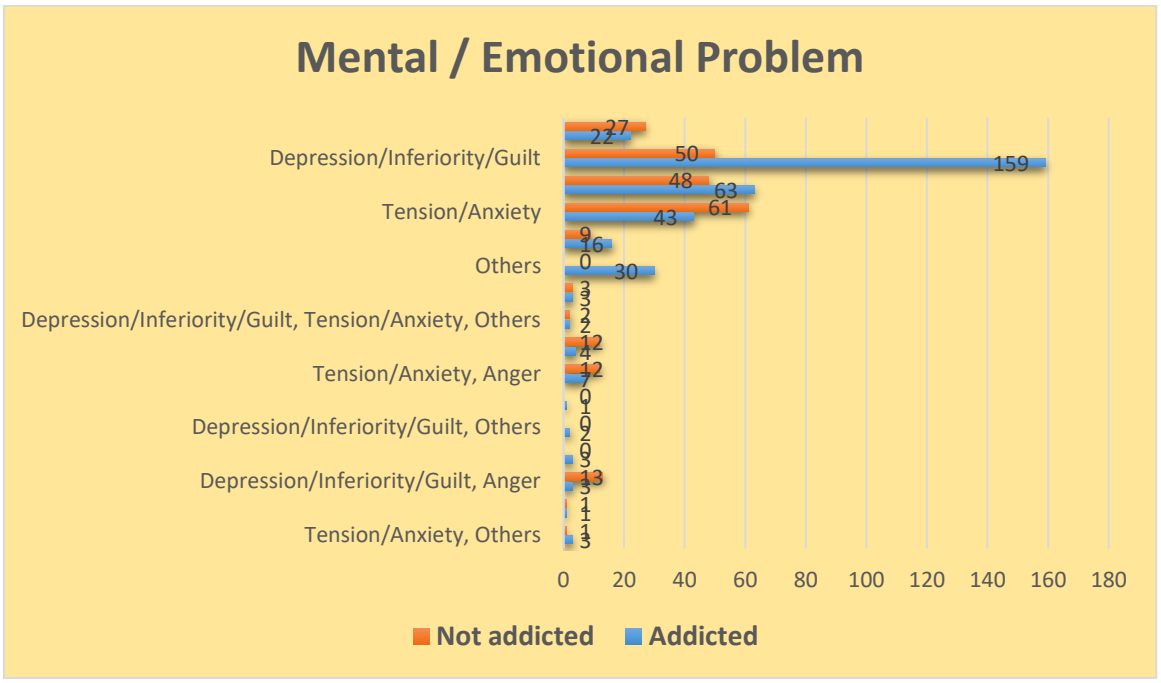

Figure 3.9: Mental / Emotional problem

Figure 3.10 is based on drug addicts only. In the figure below, we see that the financial standing of most of the families of drug addicts is strong. However, those with medium financial status are also more prone to addiction.

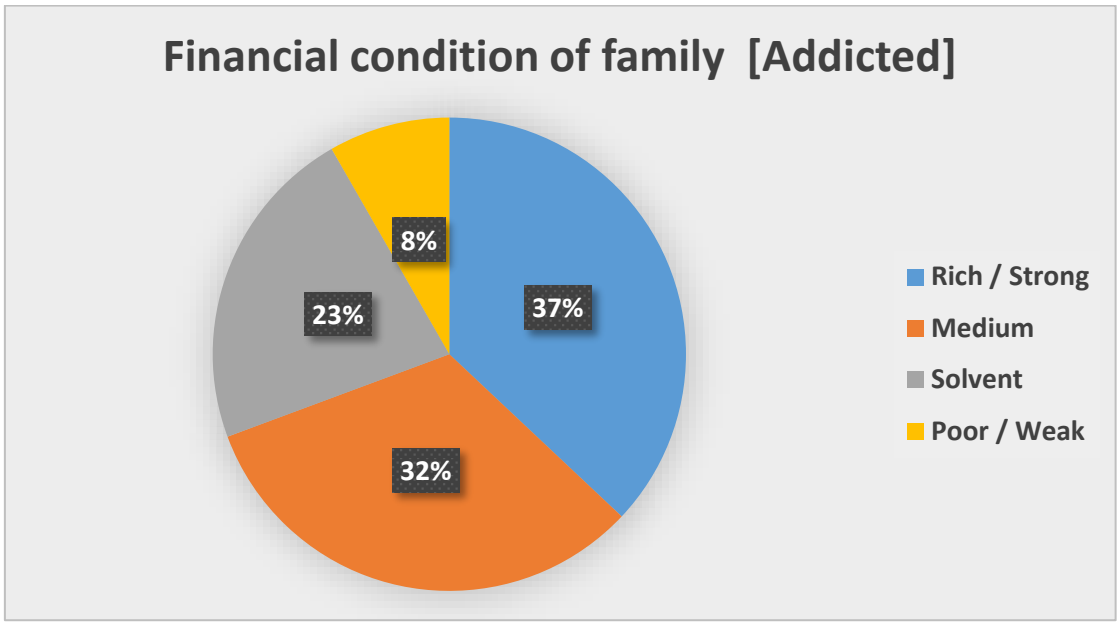

Figure 3.10: Financial condition of family vs Addicted case

It is said that addiction starts with smoking. From Figure 3.11, 80% of drug users smoke every day. On the other hand, the proportion of addicts among non-smokers is less than 5%.

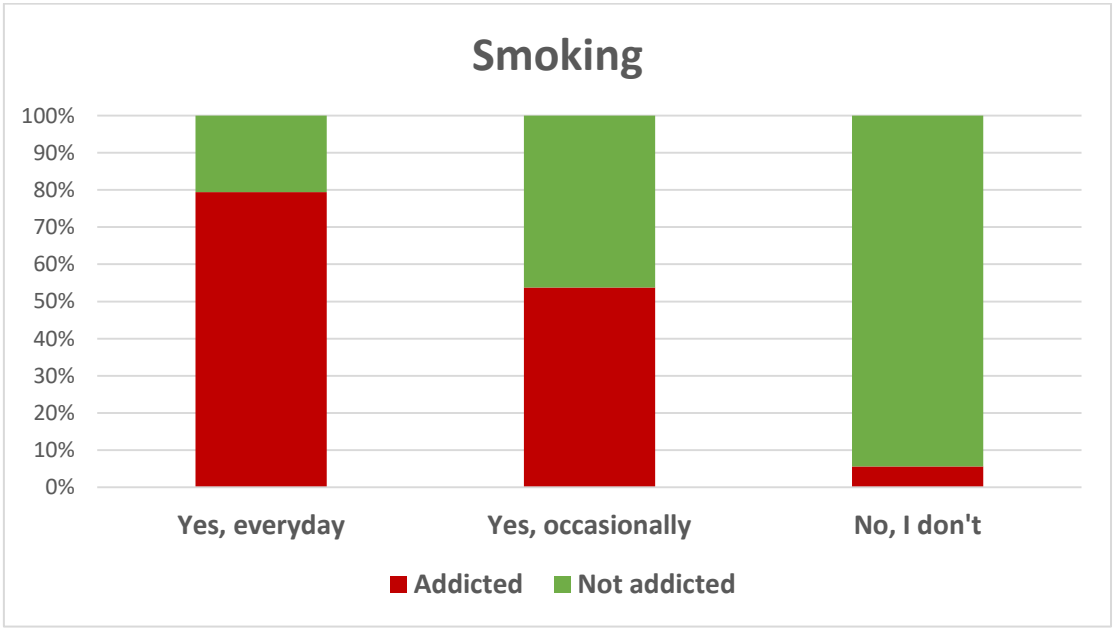

Figure 3.11: Smoking vs Addiction

Staying at a friend's house at night is a significant risk factor for drug addiction. As Figure 3.12 shows that those who often and sometimes stay at a friend's house at night are the most addicted.

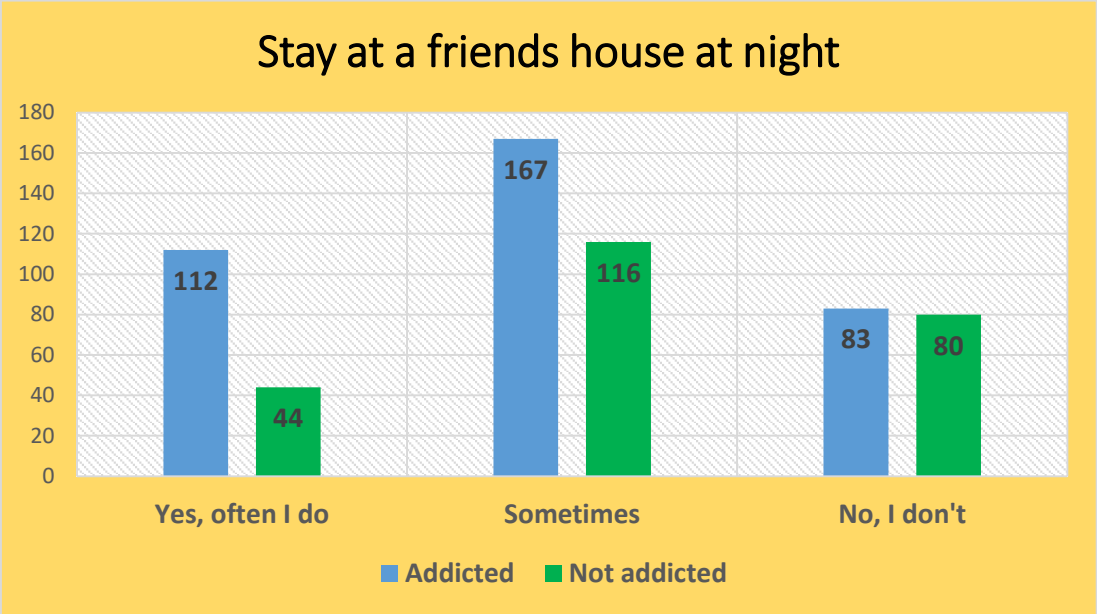

Figure 3.12: Staying at a friend's house at night

Figure 3.13 represents those who have addicted people in their family from both addicted and non-addicted data. Here we see that those who have an addicted person in their family are more likely to become addicted than not addicted.

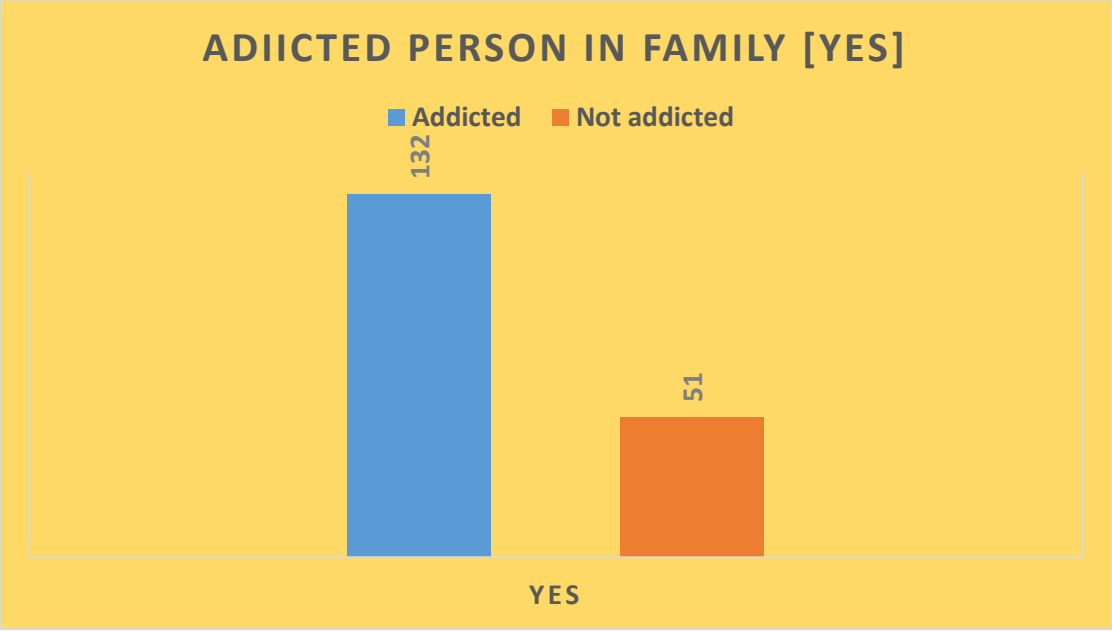

Figure 3.13: Addicted person in family vs Addiction

Figure 3.14 shows the data of those who are not satisfied with their work or education. Among them who are dissatisfied with their workplace, addicts are greater than non-addicts.

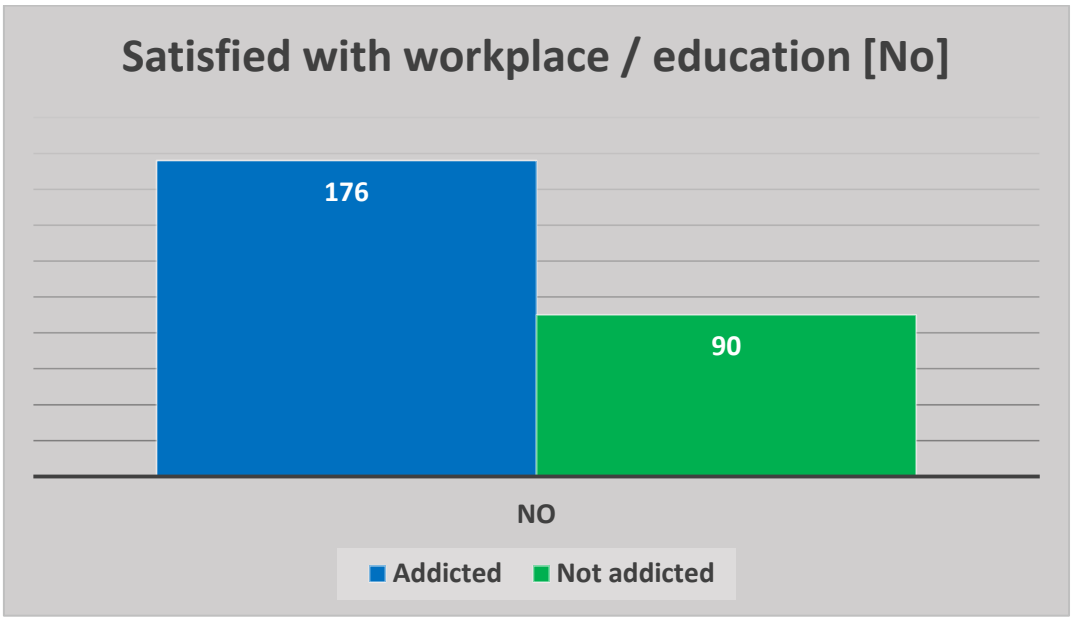

Figure 3.14: Dissatisfied with workplace vs Addiction

Figure 3.15 is based on individuals who had suicidal thoughts. Among them, most drug addicts had suicidal thoughts.

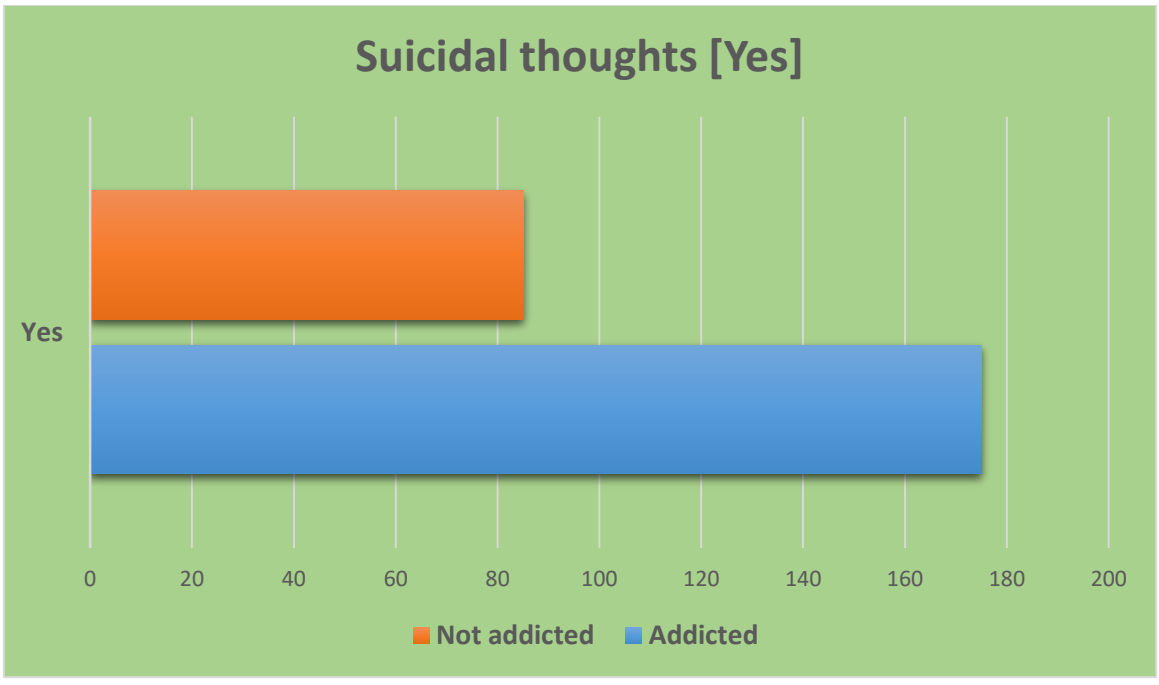

Figure 3.15: Suicidal thoughts vs Addiction

Many people become addicted to drugs because of living with drug users and a habit of smoking. Those who take drugs once are more likely to become addicted, and those who are addicted are more likely to have withdrawal symptoms. Figure 3.16 shows the correlation matrix, what kind of data we had in our data set, and how they relate to our result or outcome.

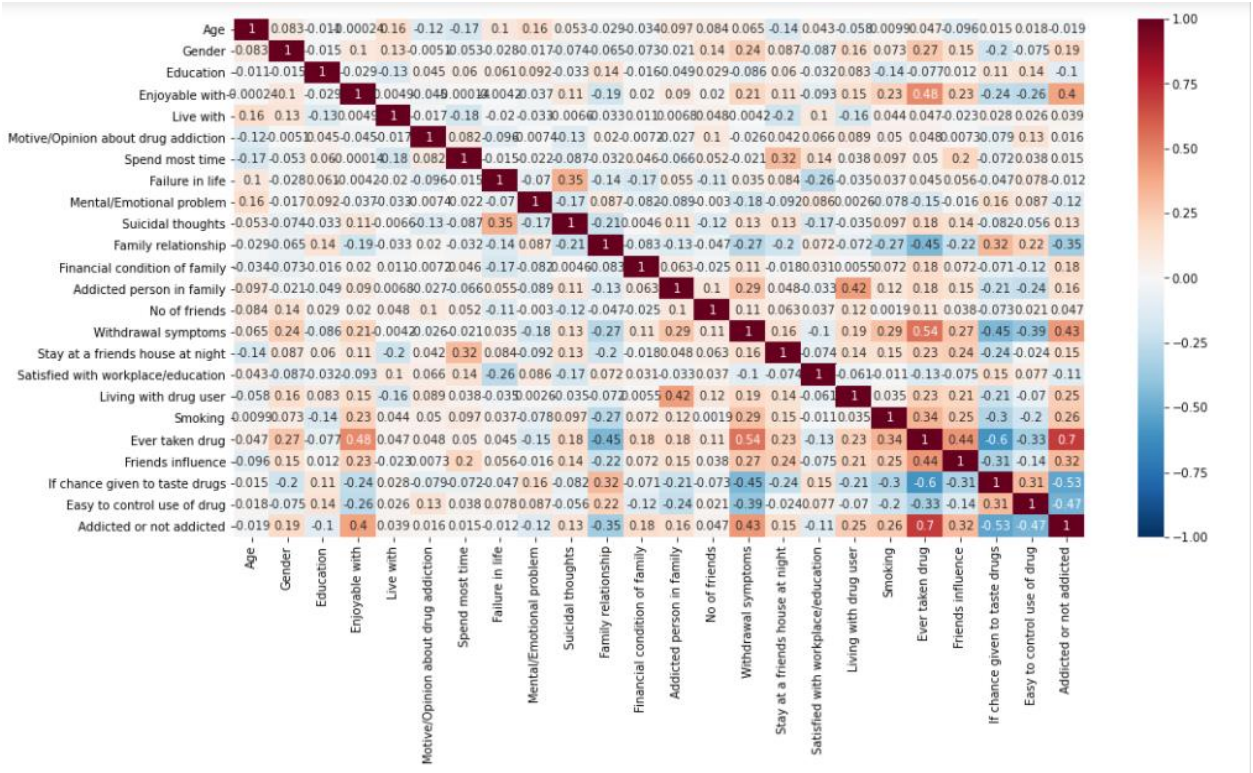

Figure 3.16: Correlation Matrix

3.6 Implementation Requirements

We need data mining tools to implement our work, data storing tools for data storage, and data processing tools for data preprocessing. We collect our data through Google Forms and handwritten forms. We have enlisted the help of Microsoft Excel to create the dataset. Our dataset file was created in CSV format. We use "Anaconda navigator" and "Jupyter notebook" for data preprocessing and algorithm implementation.

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution. Navigator allows us to launch common Python programs and easily manage conda packages, environments, and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository. [20]

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

In the previous chapter, we discussed the dataset and its preprocessing processes for our proposed model. We talked in detail about the data cleaning process, and we displayed some significant features in distinction to our target. We have used processed data in some algorithms and the results of those algorithms will be discussed in this chapter. After cleaning the data, there were 20 features in our dataset. Of the 602 data samples, 80 percent were used for training purposes and 20 percent for testing. To get the accuracy and result of our proposed model, we executed various machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, SVM, and K-Nearest Neighbor. The name of our dataset is "Drug Addiction v1".

4.2 Experimental Results and Analysis

We execute six machine learning algorithms on our processed data and compared them by calculating their accuracy, confusion matrix, precision, recall, and F1 score.

4.2.1 Experimental Evaluation

Figure 4.1 shows the accuracy of six algorithms. In this figure, the green bar represents the accuracy of all algorithms we used. Here we can see all the six classifiers gave a very close result except Decision Tree. Logistic Regression has achieved 88.42% accuracy, Decision Tree has achieved 77.68% accuracy, Random Forest has achieved 85.12% accuracy, Naïve Bayes has achieved the best accuracy of 90.90%, SVM has achieved 85.95% accuracy and kNN has achieved the accuracy of 83.47%. Among the six algorithms, we have received the highest accuracy of 90.90% in Gaussian Naïve Bayes and the lowest accuracy of 77.68% from the Decision Tree classifier.

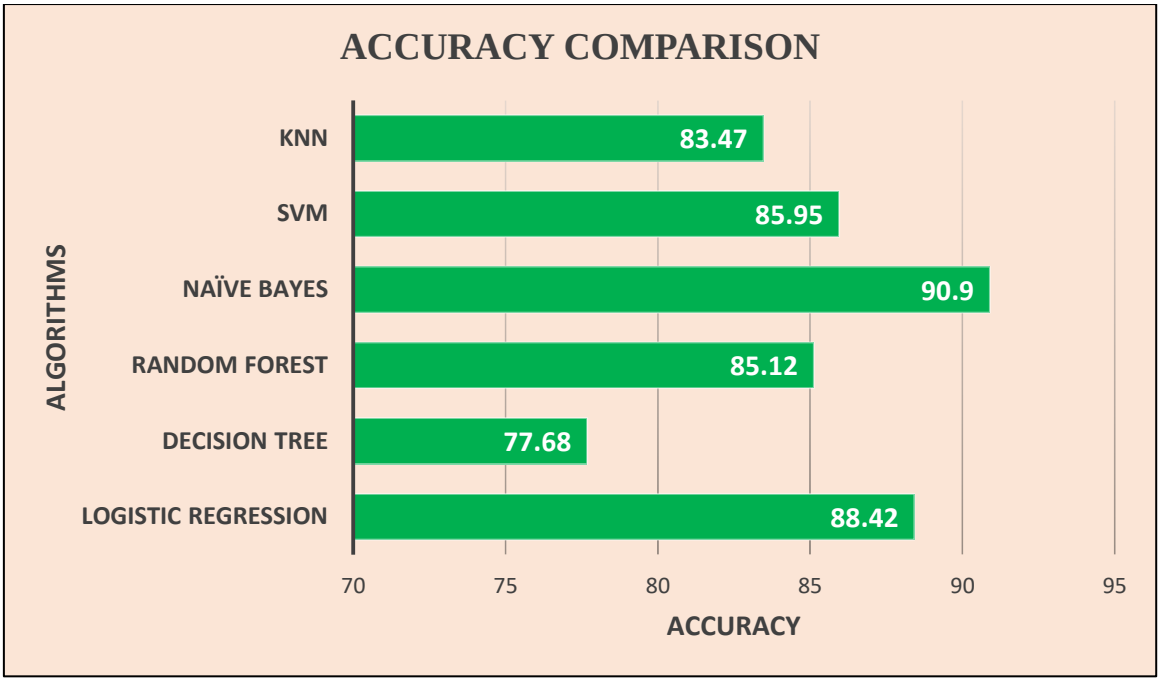

Figure 4.1: Accuracy comparison between algorithms

Logistic regression is a widely used method for classifying data under supervised machine learning. It is best to use logistic regression when the nature of the target variable is dichotomous and needs probabilistic results, and also needs to understand the impact of the feature. The logistic regression uses the sigmoid function and gives the probability of a certain class of data that has an interval of 0, 1 [22].

The Decision Tree algorithm is used to establish classification models in the shape of a tree-like structure, just like its name and it belongs to supervised class learning. It splits the dataset into smaller subsets and presents them in the manner of a tree node. In the decision tree, internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome. It is easier to understand as it follows the similar process which a human follows when making any decision in real life [21].

Random forest assures a more exact result with a large number of data items and decisions than a single Decision tree does. The algorithm generates many decision trees and then gets the prediction from each of them and finally predicts by taking the average or mean of the output from various trees. It can make a reasonable prediction without hyper-parameter tuning and possess very high accuracy [21].

Naive Bayes is a probabilistic machine learning model that is used for classification techniques based on Bayes Theorem. The word naive is used because it considers that all the predictors are independent of each other. In other words, the assumption is that the existence of a feature in a class is independent of the existence of any other feature in the same class. It can be simply coded and predictions can be made quickly in practice. Therefore, it is the preferred algorithm for smoothly scalable and traditionally real-world apps that are required to respond fast to user requests [21].

Support Vector Machine is a machine learning algorithm that is used in classification in most of the cases of the research field. The goal of the SVM algorithm is to construct optimal lines or decision boundaries that can separate n-dimensional space into classes so that we can place new data points in the appropriate categories in the future. The boundary of this best decision is called a hyperplane of SVM. This algorithm is most efficient on datasets with multiple features such as financial or medical data [22].

K-Nearest Neighbor is a versatile machine learning algorithm that falls under the supervised learning category. It is used to reset missing values and datasets and is usually used for classification purposes. It estimates the probability that a data point will become part of a group or another group based on what group is the closest to data points. In another word, it aims to put in all of the nearest neighbors around a new anonymous data point to determine what class it belongs to. It is a versatile algorithm that is very easy to interpret and can perform perfectly with sufficient data [22].

4.2.2 Descriptive Analysis

Accuracy is not the only metric to evaluate a model. That is why we also find out the precision, recall, F1 measure, and confusion matrix for each classifier.

Let us introduce the precision first,

Precision speaks about how accurate the model is between total predicted positive observations and the actual positive observations. It helps to determine when the costs of False Positives are high.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \times 100\% \qquad (1)$$

Recall truly counts how many of the Actual Positives our model capture through labeling it as Positive. Recall becomes high only when the ratio between true positive and the sum of true positive and false negative are equal.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \times 100\% \qquad (2)$$

F1 is a good measure if we need to look for a balance between Precision and Recall as it is the weighted average of precision and recall. It is more useful if we have an uneven class allocation. It becomes high when the score of precision and recall is high and it becomes low when both scores are low.

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \qquad (3)$$

The following figure 4.2 shows the comparison of Precision, Recall, and F1 measures for all the algorithms we applied. Here we can see, from Logistic Regression we got 85% precision, 97% recall, and the f1 measure is 91%. Decision Tree has achieved 84% precision, 78% recall, and 81% f1 measurement. Random Forest gave us 86% precision, 89% recall, and 88% f1 measure. Naïve Bayes has achieved 88% precision, 99% recall, and 93% f1 score. SVM scores 82% precision, 97% recall and 89% f1 measure. kNN gave us 82% precision, 93% recall, and 87% f1 score.

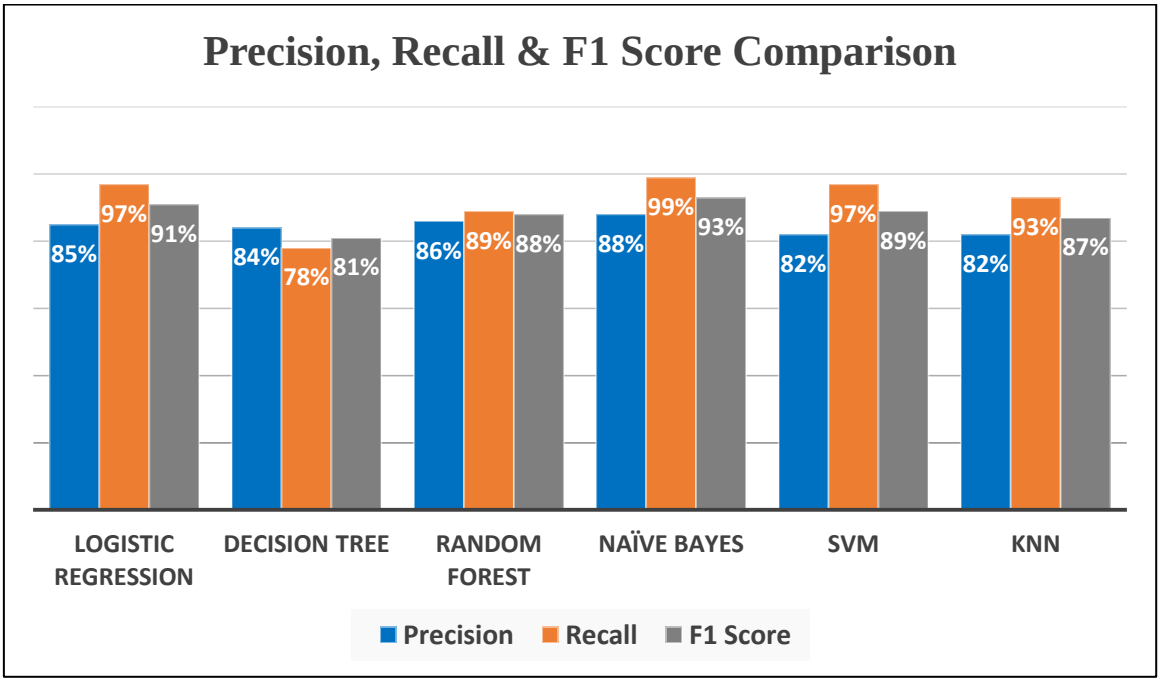

Figure 4.2: Comparison of Precision, Recall, and F1 Score

According to the comparison of Precision, Recall, and F1 measurements for all classifiers, Gaussian Naïve Bayes algorithm has achieved the highest result.

Confusion matrix comes into the limelight where a model needs to evaluate based on performance & effectiveness. Computing a confusion matrix is a much better way that can give us a better result of what our classification model is getting right and what kinds of errors it is making. In the confusion matrix, each row represents an actual class and each column represents a predicted class. Table 4.1 describes the confusion matrix of each algorithm we used in our research.

TABLE 4.1: CONFUSION MATRIX OF ALL ALGORITHMS

Algorithm	Confusion Matrix		Actual	
			No	Yes
Logistic Regression	Predicted	No	37	12
		Yes	2	70
Algorithm	Confusion Matrix		Actual	
			No	Yes
Decision Tree	Predicted	No	38	11
		Yes	16	56
Algorithm	Confusion Matrix		Actual	
			No	Yes
Random Forest	Predicted	No	39	10
		Yes	8	64
Algorithm	Confusion Matrix		Actual	
			No	Yes
Gaussian Naïve Bayes	Predicted	No	39	10
		Yes	1	71
Algorithm	Confusion Matrix		Actual	
			No	Yes
SVM	Predicted	No	34	15
		Yes	2	70
Algorithm	Confusion Matrix		Actual	
			No	Yes
k-Nearest Neighbor	Predicted	No	34	15
		Yes	5	67

Once we have built our model, the most significant question that arises is how good is our model? So, evaluating our model is the most essential task. Consequently, we evaluate our model based on Accuracy, Precision, Recall, F1 Score, and Confusion Matrix. The following Table 4.2 presents the Accuracy, Precision, Recall, and F1 Score for each of the algorithms for the sake of our research.

TABLE 4.2: ALGORITHM PERFORMANCE EVALUATION

Algorithms	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
Logistic Regression	88.42	85.36585	97.22222	90.90908
Decision Tree	77.68	83.58208	77.77777	80.57553
Random Forest	85.12	86.48648	88.88888	87.67122
Naïve Bayes	90.90	87.65432	98.61111	92.81045
SVM	85.95	82.35294	97.22222	89.17197
kNN	83.47	81.70731	93.05555	87.01298

4.3 Discussion

This chapter reviews the performance of all six algorithms based on accuracy, precision, recall, and F1 score. Besides, discussed the function and equation of the evaluation matrix and models. From the performance of all algorithms, we can see that Gaussian Naïve Bayes delivers the highest accuracy of 90.90%. Also, Gaussian Naïve Bayes achieved 87.65% precision, 98.61% recall, and 92.81% F1 score. Therefore to conclude, we find out that, using the Naive Bayes algorithm, we get the best performance of our model for the prediction of being addicted to drugs.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Findings and Contributions

The goal of the study was to create a model for predicting the risk of any individual towards drug addiction. Therefore, first, we collected data from FERA Rehabilitation Center & AMI Addiction Management Institute. Furthermore, we collected data from the Kaggle community and colleges, universities. There were 60% drug addicts & 40% non-addicts information in our dataset. To get a clean & pure dataset, we do data preprocessing on our raw dataset by using the Anaconda navigator & Jupyter Notebook. Since one of the objectives of this study was to dig into the root causes behind drug addiction, so, using feature selection techniques, we find the most influential factors behind addiction. As a result, we drop the less important feature for our study. After getting the processed dataset, we used six popular machine learning algorithms for our proposed model. Among these algorithms, we selected our model based on their performance measurements which include, accuracy, precision, recall, and F1 score. We found that Naive Bayes gave the best results with maximum accuracy, precision, recall, and F1 score. Therefore, we selected the Gaussian Naïve Bayes algorithm as our model to predict the risk of an individual's drug addiction.

5.2 Recommendations for Future Work

Our research work has some limitations and shortcomings. Due to some limitations, we were unable to include data from all districts of Bangladesh in our study. In addition, more advanced methods of machine learning might have been applied, to our research, in which case, the research would have been better presented, and the results of the research would have been different. In the future, performing new algorithms, counting various parameters, and adding more features can be made it a more effective model. Also, a larger and more powerful dataset can be created by collecting data from people in all districts. In addition, the model can be enlarged and improved with the help of the Department of Narcotics Control.

REFERENCES

[1]. Arifur Rahman Rabbi, "43% of unemployed population addicted to drugs", Dhaka Tribune, 2019. [Online]. Available: https://www.dhakatribune.com/bangladesh/dhaka/2019/02/27/43-of-unemployed-population-addicted-to-drugs. [Accessed: 25- June- 2021].

[2]. "RAB arrests 4 with hallucinogenic drugs LSD, DMT", Dhaka Tribune, 2021. [Online]. Available: https://www.dhakatribune.com/bangladesh/crime/2021/06/27/rab-arrests-4-with-hallucinogenic-drugs-lsd-dmt. [Accessed: 01- July- 2021].

[3]. Shaheen Mollah, "Restricted, she killed parents", The Daily Star, 2013. [Online]. Available: https://www.thedailystar.net/news/restricted-she-killed-parents. [Accessed: 01- July- 2021].

[4] A. A. Choudhury, Md. R. H. Khan, N. Z. Nahim. et al. "Predicting Depression in Bangladeshi Undergraduates using Machine Learning," 2019 IEEE Region 10 Symposium (TENSYMP 2019), PID: 306, Page: 789.

[5] A. Amirabadizadeh, H. Nezami, M. G. Vaughn et al. "Identifying Risk Factors for Drug Use in an Iranian Treatment Sample: A Prediction Approach Using Decision Trees," Substance Use & Misuse (2017). doi: 10.1080/10826084.2017.1392981.

[6] J. Choi, Hee-Tae Jung, A. Ferrell et al. "Machine Learning-Based Nicotine Addiction Prediction Models for Youth E-Cigarette and Waterpipe (Hookah) Users," J. Clin. Med. 2021, 10, 972. https://doi.org/10.3390/jcm10050972.

[7] YoungJin Choi, YooKyung Boo et al. "Comparing Logistic Regression Models with Alternative Machine Learning Methods to Predict the Risk of Drug Intoxication Mortality," Int. J. Environ. Res. Public Health 2020, 17, 897.

[8] Mohaamad H Afzali, Matthew Sunderland, Sherry Stewart et al. "Machine-Learning Prediction of Adolescent Alcohol Use: A Cross-Study, Cross-Cultural Validation". doi: 10.1111/add.14504.

[9]. Y. Zhang, J. Liu, Z. Zhang and J. Huang, "Prediction of Daily Smoking Behavior Based on Decision Tree Machine Learning Algorithm," 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 2019, pp. 330-333.

[10]. D. Kumari, S. Kilam, P. Nath. et al. "Prediction of alcohol abused individuals using artificial neural network". Int. j. inf. tecnol. 10, 233–237 (2018).

[11]. S. Kinreich, J. L. Meyers, A. Maron-Katz. et al. "Predicting risk for Alcohol Use Disorder using longitudinal data with multimodal biomarkers and family history: a machine learning study". Mol Psychiatry (2019).

[12]. D. Dahiwade, G. Patle and E. Meshram, "Designing Disease Prediction Model Using Machine Learning Approach," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 1211-1215.

[13]. A. M. Alaal, T. Bolton, E. D. Angelantonio, J. H. F. Rudd, M. van der Schaar. "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants", PLoS One. 2019; 14(5) e0213653. doi:10.1371/journal.pone.0213653. PMID: 31091238; PMCID: PMC6519796.

[14]. Mostafa Mohiuddin Jalal Protik, "Drug Addiction in Bangladesh- Reasons". [Online]. Available: https://www.kaggle.com/protikmostafa/drug-addiction-in-bangladesh-reasons. [Accessed: 02- July- 2021].

[15]. "What Is Drug Addiction?" 2019. [Online]. Available: https://www.webmd.com/mental-health/addiction/drug-abuse-addiction#2. [Accessed: 05- July- 2021].

[16]. "Teen Drug Abuse and Recovery." 2019. [Online]. Available: https://www.nextgenerationvillage.com/drugs/. [Accessed: 08- July- 2021].

[17]. "10 Reasons Why People Abuse Drugs." 2019. [Online]. Available: https://www.recoveryconnection.com/10-reasons-people-abuse-drugs/. [Accessed: 08- July- 2021].

[18]. "Causes of Drug Addiction - What Causes Drug Addiction?" 2019. [Online]. Available: https://www.healthyplace.com/addictions/drug-addiction/causes-of-drug-addiction-what-causes-drug%20addiction. [Accessed: 08- July- 2021].

[19]. "The Causes and Effects of Drug Addiction." 2019. [Online]. Available: https://www.altamirarecovery.com/drug-addiction/causes-effects-drug-addiction/. [Accessed: 08- July- 2021].

[20]. "Anaconda Navigator." 2020. [Online]. Available: https://docs.anaconda.com/anaconda/navigator/ [Accessed: 15- July- 2021].

[21]. Jiawei Han, Micheline Kamber, Jian Pei, Data Mining Concept and Technique, 3rd Edition, Morgan Kaufmann, 2012,pp. 332-398.

[22]. Stuart J.Russell, Peter Norvig, Artificial Intelligence a Modern Approach, 3rd Edition, Upper Saddle River, NJ : Prentice Hall, 2010,pp. 725-744.

PLAGIARISM REPORT

Turnitin Originality Report

Processed on: 23-Jan-2022 16:10 +06
ID: 1746273672
Word Count: 9155
Submitted: 1

181-35-2400 By Md Tasnim Alam Fahim

Similarity Index	Similarity by Source
29%	Internet Sources: 26% Publications: 17% Student Papers: 13%

5% match (Internet from 01-Oct-2021)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5469/163-15-8432%20%2829_%29.pdf?isAllowed=y&sequence=1

2% match (student papers from 30-Nov-2021)
Submitted to Daffodil International University on 2021-11-30

2% match (student papers from 12-Jan-2021)
Submitted to Daffodil International University on 2021-01-12

2% match ()
YoungJin Choi, YooKyung Boo. "Comparing Logistic Regression Models with Alternative Machine Learning Methods to Predict the Risk of Drug Intoxication Mortality", International Journal of Environmental Research and Public Health

1% match (Internet from 05-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5695/171-35-1900%20%2817_%29.pdf?isAllowed=y&sequence=1

1% match (Internet from 05-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5685/171-35-1838%20%2813_%29.pdf?isAllowed=y&sequence=1

1% match (Internet from 15-Mar-2020)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/3553/P13659%20%2829%25%29.pdf?isAllowed=y&sequence=1

1% match (Internet from 05-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5692/171-35-1862%20%283_%29.pdf?isAllowed=y&sequence=1

1% match (Internet from 26-Feb-2020)
https://onlinelibrary.wiley.com/doi/abs/10.1111/add.14504

1% match (Internet from 02-Dec-2021)
https://pdfs.semanticscholar.org/889e/b11874790cfbbf235201f6c19f2c98b492fe.pdf

1% match (publications)
Ahnaf Atef Choudhury, Md. Rezwan Hassan Khan, Nabuat Zaman Nahim, Sadid Rafsun Tulon, Samiul Islam, Amitabha Chakrabarty. "Predicting Depression in Bangladeshi Undergraduates using Machine Learning", 2019 IEEE Region 10 Symposium (TENSYMP), 2019

1% match (Internet from 13-Nov-2021)
https://www.science.gov/topicpages/d/decision+trees+dt

1% match (Internet from 17-Oct-2019)
https://link.springer.com/article/10.1007%2Fs41870-018-0094-3

1% match (Internet from 20-Jan-2022)
https://www.educative.io/edpresso/what-is-anaconda-navigator

1% match (Internet from 23-Oct-2020)
https://www.nature.com/articles/s41380-019-0534-x?code=b195e35d-4dd1-43e6-a725-9bc5c1f427f6&error=cookies_not_supported

< 1% match (Internet from 05-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5733/163-35-1778%20%2814_%29.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 05-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5683/171-35-1827%20%2819_%29.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 09-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/6655/172-15-9705%20%2822%25%29.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 05-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5693/171-35-1870%20%2824_%29.pdf?isAllowed=y&sequence=1