Predicting Drug Addiction Patterns in Urbanizing Bangladesh: A Machine Learning Approach

By

Md.Khaledur Rahman Onik
191-15-12959

Nakib Hossen Naim
201-15-14175

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Dr.S.M. Aminul Haque
Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Mr.Mehadi Hasan
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

January 12, 2025

APPROVAL

This Project titled **"Predicting Drug Addiction Patterns in Urbanizing Bangladesh: A Machine Learning Approach"**, submitted by **Md. Khaledur Rahman Onik** and **Nakib Hossen Naim**, ID No: **191-15-12959,201-15-14175** and to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 12 January, 2025.

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Sharmin Akter — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Md Mohammad Masum Bakaul — **Internal Examiner**
Sr. Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud — **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S.M. Aminul Haque**, **Professor and Associate Head** , Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. S.M. Aminul Haque
Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

11.01.2025

Mr. Mehadi Hasan
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Onik

Md. Khaledur Rahman Onik
Student ID: 191-15-12959
Department of Computer Science and Engineering
Daffodil International University

Nakib

Nakib Hossen Naim
Student ID: 201-15-14175
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project(FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Dr. S.M. Aminul Haque**, **Professor and Associate Head**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Machine Learning and Data Analysis** 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 the **Dr. Sheak Rashed Haider Noori, Professor and Head,** Department of CSE Daffodil International University, Dhaka, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

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

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

ABSTRACT

Drug addiction is a significant public health challenge worldwide, with its impact on individuals, families, and communities increasing in fast urbanizing countries. In Bangladesh, the rapid growth of cities, socio-economic inequities, cultural shifts, and increased drug accessibility have led to complicated addiction patterns that require quick treatment. Factors such as peer pressure, unemployment, mental health issues, and family history of substance use play crucial roles in the onset and progression of addiction. Machine learning technologies have emerged as strong tools in forecasting complicated human behaviors, including addiction risks. This study uses Logistic Regression, Decision Trees, Random Forest, Support Vector Machine, Naive Bayes, AdaBoost, KNN, and MLPClassifier to find significant factors of drug addiction in Bangladesh and construct a risk prediction model. This model assesses the probability of addiction at individual and community levels, providing data for politicians and healthcare practitioners to plan focused therapies. these are result of ten models. Logistic Regression result is 0.80, Decision Tree result is 0.77, Random Forest result is 0.79, SVM result is 0.80, Naive Bayes result is 0.80, AdaBoost result is 0.81, KNN result is 0.77, MLPClassifier result is 0.79, Gradient Boosting result is 0.80, CatBoost result is 0.81This research fills a vital information gap by blending demographic, behavioral, and familial data to offer actionable insights. By adopting a data-driven approach, this study aims to provide a deeper knowledge of the distinct risk profiles of individuals and communities in Bangladesh. The use of predictive modeling methods can offer a more exact way to quantify addiction risks, enabling healthcare practitioners and governments to spend resources more effectively and devise focused therapies. The validated predictive model will evaluate the chance of drug addiction for individuals in Bangladesh, offering a risk score ranging from 0 to 100% and dividing individuals into low-risk and high-risk groups based on important demographic, psychological, and environmental factors. The findings will enable the development of intervention programs and evidence-based policy recommendations aimed at lowering addiction risks. In conclusion, this study serves as a significant step in addressing drug addiction in Bangladesh, producing a healthier and more supportive society.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Drug addiction presents an important public health challenge worldwide, with its impact on individuals, families, and communities typically increased in fast urbanizing countries. In Bangladesh, the increased speed of urbanization has brought both benefits and risks, with one of the main issues being increasing levels of substance abuse. As cities expand, socio-economic inequities, shifts in cultural values, and increased accessibility to drugs lead to complicated addiction patterns that necessitate quick treatment. The impact of drug addiction in Bangladesh are far-reaching, that extend economic cost, social stigma, and a pressure on public health services. Studies from urbanizing countries suggest that demographic, psychological, and environmental factors such as peer pressure, unemployment, mental health issues, and family history of substance use play crucial roles in the onset and progression of addiction [1] [2]. Furthermore, addiction risk often changes between areas, age groups, and socioeconomic classes, suggesting created therapies. Machine learning technologies have emerged as strong tools in forecasting complicated human behaviors, including addiction risks. We use this model Logistic Regression, Decision Trees, Random Forest, Support Vector Machine, Naive Bayes, AdaBoost, KNN, MLPClassifier. By employing ML, this work intends to find significant factors of drug addiction in Bangladesh and construct a viable risk prediction model. This model tries to assess the probability of addiction at individual and community levels, giving significant data for politicians and healthcare practitioners to plan focused therapies. Given the scarcity of research on addiction modeling in developing nations, notably in Bangladesh, this work fills a vital information gap. It blends demographic, behavioral, and familial data to offer actionable insights [3]. Through a thorough examination of quantitative and qualitative characteristics, this research intends to contribute to the global debate on addiction while addressing the particular socio-cultural dynamics of Bangladesh

1.2 Motivation

Bangladesh has seen the alarming growth in drug addiction in urbanising areas, a problem that requires evidence based therapy. The main socio-economic issues which also contribute to the rising prevalence of substance misuse typically associated with urbanization are income disbalance, shifts of culture and the increased exposure of risk factors like peer pressure and stress, etc. . Though it can impact public health and so few studies seek to predict modeling how drug addiction will impact individuals, policymakers and health care professionals can not build concentrated treatments without this data.Machine learning can help fill this gap by looking at several parameters of the data to identify risk factors and forecast a person's probability of being hooked on drugs. Such models are specifically important in culture and social context such as Bangladesh, where addiction is usually marginalised and intervention opportunities are small . It is possible to detect at-risk individuals and groups through data driven insight, making it possible to transform the tactics of 'prevention' and to reallocate resources.This motivated research aims at combining the need to understand addiction tendencies in an ever changing society with gaining new statistical methods for preventing the spread of addiction. This study strives to formulate a mechanism that closes the gap between availability of data and working solutions to existing problems and to use this to devise sustainable public health measures with a focus on overcoming the rising problem of drug addiction in Bangladesh, especially in urban settings. Consequently, there is an increasing problem in polydrug usage, a problem tied to new social, economic, and environmental factors in the country as urbanization accelerates . Attempts to address this problem have proven unsuccessful with a surplus of incomplete models that claim to predict and manage the risk factors of addiction but are actually clearly inadequate. Solving this gap, this work builds a statistical model capable of identifying important predictors for addiction and provides actionable insights that can guide intervention strategies.

1.3 Objectives

The research is expected to produce a validated predictive model capable of evaluating the chance of drug addiction for individuals in Bangladesh. This model will offer a risk score ranging from 0 to 100% and divide individuals into low-risk and high-risk groups based on important demographic, psychological, and environmental factors. Additionally, the study will uncover the most critical elements contributing to addiction, such as peer pressure, family history, urbanization, and socio-economic disparities, bringing important understandings into the underlying causes of substance misuse.Ultimately, the research will serve as a key tool for healthcare providers, legislators, and community organizations to confront the growing burden of drug addiction in a fast-urbanizing Bangladesh..

1.4 Methodology

Data is gathered through survey utilizing both online and offline sources and then data is prepared to decrease bias and build quality. This is because features engineering is needed to answer a number of questions the survey can answer. We use already processed datasets to apply and train machine learning techniques including LR, DT, RF, Naive Bayes, KNN, SVM, and SGD. The end results of the project are presented followed by the results assessed using F1 score, recall, accuracy, and precision. The top performing model is determined by Accuracy, Scalability and Computing efficiency.A front end website is designed in HTML, CSS, JavaScript along with some frameworks, and the express framework is used to provide backend to users for easy interaction. several questions, features engineering is essential. Datasets that have already been processed are used to apply and train machine learning algorithms such as LR, DT, RF, Naive Bayes, KNN, SVM, and SGD. Results are presented at the end of the project and assessed using metrics such as F1-score, recall, accuracy, and precision. Accuracy, scalability, and computing efficiency are the criteria used to choose the top-performing model. HTML, CSS, JavaScript, and other frameworks are used to design a front-end website, and the express framework is used to create a backend for user-friendly interaction.

1.5 Project Outcome

It is expected that this research will lead to a model that can predict the likelihood of developing drug addiction for people in Bangladesh. Here, we create a model which will provide a risk score between 0 and 100% of an individual's risk based on major demographic, psychological & environmental factors and then separate individuals into low and high risk groups. The study will also discover what key matters to the addiction of the people especially whom are affected with such problem, such as pressure from the others, and family history, urbanization, and socio-economic differences, giving more understanding in the main causes of substance misuse. The findings will ensure the development of created intervention programs as well as evidence based policy recommendations for reduction of addiction risk. However, this research would act as an important tool for healthcare providers, legislators and community organizations to address the dramatic rise in drug addiction in a fast urbanising Bangladesh.

1.6 Organization of the Report

The report is intended to provide a clear and thorough narrative of the study method, findings, and consequences. Each chapter serves a specific purpose, ensuring cohesion and depth in presenting the research.
Chapter 1: Introduction This chapter includes an overview of the report. It addresses the general overview of the topic of concern before proceeding to highlight on the rationale of

the project. The purpose of this section is to provide an overview of the research goals of this study, the approach used in the study, the anticipated project's outcomes, and the layout of the report as a useful aid to readers.

Chapter 2: Background This chapter forms the background and context of the research. It consists of an overview of the background and development of the particular research, secondary to the literature review examining previously conducted research in the subject area. This study need is further established given that a gap analysis is made in order to point out where previous studies lack. The last section of the chapter presents the conclusion of the study.

Chapter 3: Research Methodology This chapter presents details regarding the approach used in the project. This writing commences with a brief description of the research methodology and the general plan of the system's design; it also provides a comprehensive description of the steps involved in the actualization of the proposed research methodology and system design. The chapter also contains the project plan, tasks to be assigned and the chapter summary.

Chapter 4: Implementation and Results This chapter is devoted to the description of the key practical aspects of the work. They include setting up of the environment, testing as well as evaluating the system. A performance analysis is presented, as well as a comparative analysis. The findings are further explored while placing critical focus on their status and relation to objectives. The conclusion of the chapter is a review of the findings of the study.

Chapter 5: Engineering Standards and Design Challenges This chapter focuses on the engineering regulations of the project such as software and hardware engineering regulations. They articulate the social relation of the project to society and the environment, ethical question, and sustainability measures. Third, it measures the efficiency of project management and assesses the financial aspect of the project and offers an approach towards solving engineering issues for conduction the job. The last section of the chapter is a conclusion.

Chapter 6: Conclusion Thus, in the last chapter, the author articulates conclusion that in essence reflects all the findings and accomplishments of the report. The report is closed out with a discussion on the limitations and a listing of potential future work that can be done on the study.

Chapter 2

Background

2.1 Introduction

Fundamental terminologies and concepts of predictive modeling of drug addiction in an urbanizing Bangladesh are described in this study. Drug addiction is defined as a chronic, relapsing disorder characterized by compulsive drug seeking, use despite severe consequences and changes in the brain. Urbanization is a term that denotes an increase in the number of people living in urban areas associated with socioeconomic transformation that may make the risk factor of these few elements (peer Pressure, unemployment and stress), which could lead one to addiction in the future. ML is the use of algorithms and statistical models to analyze and interpret messy datasets, and predict outcomes such as addiction rates. ML is an application of various predictive modeling approaches to predict the likelihood of future behaviors or outcomes, given existing historical data. Demographic variables such as age, income, education; psychological characteristics, such as impulsivity and propensity for modifying behavior under peers; environmental factors such as accessibility to drugs, and a family history of addiction; are also part of this study. The interpretation of results from this research depends on understanding these terminologies.

2.2 Literature Review

2.2.1 Similar Applications

Performances have used similar use and investigation of predictive modeling and machine learning to respond to problems in society such as addiction and public health. Zhao et al. showed that behavioral and demographic aspects were vital to making addiction prediction by using logistic regression and decision trees to predict risks of substance misuse among the juvenile population [1]. According to Nguyen et al. , neural network models, like LSTM, were used to model complicated patterns to predict the relapse risks in recovered addicts [2]. This study adopted a multi factorial approach similar to the case studies by Volkow et al. which highlighted the interaction within urban stressor × genetic predisposition to addiction [3]. Web based platforms such as RecoveryAI, which employ

Prediction Analytics to generate individualized addiction treatment suggestions show the potential for AI driven technologies in public health. There are ways to incorporate data driven strategies into such easy to use applications as Mobile apps like Sober Grid that provide real time support and also track user behaviour using machine learning algorithm. These studies and applications are directly complemented by the focus of this research, which show methodological advances and stress the importance of using machine learning to integrate demographic, psychological, and environmental variables for predictive and preventative treatments.

2.2.2 Related Research

The study of drug addiction by predictive modeling has been extensively examined, particularly in high-income countries, but remains under-researched in developing nations like Bangladesh. The following part covers major publications in this subject, giving a basis for understanding the present research environment and identifying gaps that this study intends to remedy. Rahman et al. investigated the socio-economic elements contributing to drug addiction in urbanizing Bangladesh, highlighting the importance of peer pressure, unemployment, and mental health difficulties [4, 5, 6]. Similarly, Khan et al. explored the socio-economic drivers of addiction in South Asia, underlining the necessity for tailored interventions [7]. Volkow et al. highlighted the interplay between urban stressors and genetic predispositions to addiction, supporting a multi-factorial model of risk [3]. Advancements in machine learning applications for addiction prediction have been studied by Zhao et al, who employed decision trees and logistic regression to predict substance misuse risk in juvenile populations [1]. Likewise, Bickel et al.explored reinforcement learning models to identify behavioral tendencies leading to addiction [8]. Ghitza et al. proved the efficacy of predictive analytics in tailoring addiction treatment regimens, underlining the need of individualized methods. In the arena of environmental factors [9] Applications of deep learning in addiction research have been rising, with Nguyen et al. building neural network models to predict relapse risks in recovered addicts [2]. Meanwhile, Kumar et al.studied the integration of socio-demographic data into machine learning frameworks, achieving good prediction accuracy in identifying at-risk individuals [10]. Recent publications have also underlined the need for explainability in addiction models. Xie et al. employed interpretable ML approaches to find significant predictors of opioid addiction, boosting the practical applicability of their models [11, 12]. Similarly, Lipton recommended for balancing model complexity with interpretability, a criterion crucial for public health applications [13, 14]. In the context of addiction patterns, Nair et al. explored regional variations in addiction prevalence, revealing insights into how socio-economic determinants intersect with geographic disparities [15, 16]. Ahmed et al.focused on the impact of family history in influencing addiction risks, finding strong correlations that guide preventive efforts [17]. Moreover, works by Zhang et al.and Lee et al. have stressed the potential of integrating qualitative data, such as self-reported experiences, with quantitative predictors to boost model robustness. These works imply that merging multiple data sources can give more thorough insights about addiction [18, 19]. In Bangladesh, Sultana et al. studied the accessibility of drugs in metropolitan areas as a crucial determinant in addiction, presenting region-specific findings [20]. While tremendous progress has been made globally, this review shows a scarcity of research concentrating on predictive modeling for drug addiction in the socio-cultural setting of Bangladesh. By addressing these shortcomings, this study attempts to contribute to the field by incorporating demographic, psychological, and environmental factors into a machine learning architecture unique to Bangladesh.

2.3 Gap Analysis

Highlights of the gap analysis table 2.1 show how the current study differ from those done previously by Rahman, Khan, Zhao, and Ghitza. Unlike the studies of Rahman, Khan, and Ghitza, this research employs primary data gathered from the target population which increases the work's applicability and specificity. The concept of personal risk factors that complement the evaluation of addiction was missing from the previous work except in Ghitza; however, it is implemented in this project to provide personalized use information. Additionally, the feature of multi-language support, which was not mentioned in all the previous studies, make it available for people speak different languages, especially where people speak different languages are common, such as Bangladesh. One especially important development is the integration into a website, where users can easily communicate with the model through an intuitive interface. This feature makes the current work different from the previous studies in which no study used an online interface for practice. Combined, these innovations help to fill previously researched shortcomings and provide a greater coverage, easier use, and more user-oriented approach to the risk assessment for drug addiction.

Table 2.1: Gap Analysis with Previous Work.

Featurs	**Rahman**	**Khan**	**Zhao**	**Ghitza**	**Our Work**
Is Primary Data	No	No	Yes	No	Yes
Personalized Risk Profiles	No	No	No	Yes	Yes
Multi-Language Support	No	No	No	No	Yes
Website Integration	No	No	No	No	Yes

2.4 Summary

It is underlined the growing number of the drug addict problem in the city where the youth is the main victims, the reason underlined is socio-economic issue especially the peer pressure, unemployment and easy availability to the narcotic. Key causes of the problem are psychological and environmental variables, and urbanization and its associated stresses. While predictive modeling in addiction research has globally succeeded, there is limited consideration for the Bangladeshi socio-cultural context. Recent studies have identified global risk indicators and employed machine learning techniques, but little study has been carried in Bangladesh. First researched by more recent works such as Rahman et al.and Sultana, et al. stresses on the features of the region that become recreative to urban stress and stigma that hamper efforts in collecting trustworthy data and giving effective solutions. Studies such as Xie et al. and Nguyen et al. have shown that studying advances in machine learning, in particular the use of explainable models, have been viable in predicting global addiction risk. But all this is very much not happening in Bangladesh. Based on these results, clear need for integrated methodologies in which demographic, psychological, and environmental aspects are in part modeled via explainable machine learning emerges. If this gap can be addressed, comprehensive, data driven strategies for preventing drug addiction, improving public health outcomes and eradicating social stigma will be achieved.

Chapter 3

Research Methodology

3.1 Methodology

3.1.1 Overview

The research leverages Google Colab for computational analysis and model construction, employing Python programming and machine learning tools. The project was separated into three phases: data collecting, data analysis including model evaluation, and deployment of the best-performing model on a web-based platform. Data gathering involved Google forms and human questionnaires, while Python tools such as Pandas and NumPy were used for preprocessing. The project was cost-efficient and scalable, with cloud-based infrastructure equipped with Tesla GPUs and ample RAM. The final deployment of the predictive models was managed using Visual Studio code, including HTML, CSS, React, Node.js, and Express.js frameworks. The project was carefully planned on a 52-week timeframe, with an estimated budget of BDT 1,37,000 to BDT 1,61,000. The structured approach of the study warrants a robust framework for predictive modeling research of drug addiction in metropolitan Bangladesh.

3.1.2 Proposed Methodology

The process begins with 3.1 data collection both online and physical survey. Data preprocessing involves removing null values and duplicate entries to enhance data quality and mitigate bias. Features Engineering is important step for this study because there are lots of question, we through the candidate during survey. In addition, various machine learning algorithms such as LR, DT, RF, Naive Bayes, KNN, SVM, SGD are applied and trained on pre-processed datasets. The figure 3.2.1 represents a typical data science project workflow. It starts with collecting survey data from Dhaka, which is then collected and formed into a dataset. This dataset undergoes pre-processing to make it clear and ready for analysis. Then, algorithms are applied to extract insights from the data. The results of these algorithms are shown for better understanding and interpretation. Finally, the project ends with the presentation of the results The model evaluation is conducted using appropriate metrics such as accuracy, precision, recall and F1-score and understand the influencing factors. The best-performing model is selected based on accuracy and other relevant metrics, taking into account computational efficiency, interpretability, and scalability. Subsequently, a front-end website is designed using HTML, CSS, JavaScript, other framework like bootstrap, Font Awesome etc. and backend developed using the express framework for intuitive and user-friendly interaction. Finally, the trained model and web application are deployed on a local server, ensuring scalability and reliability of the deployed system.

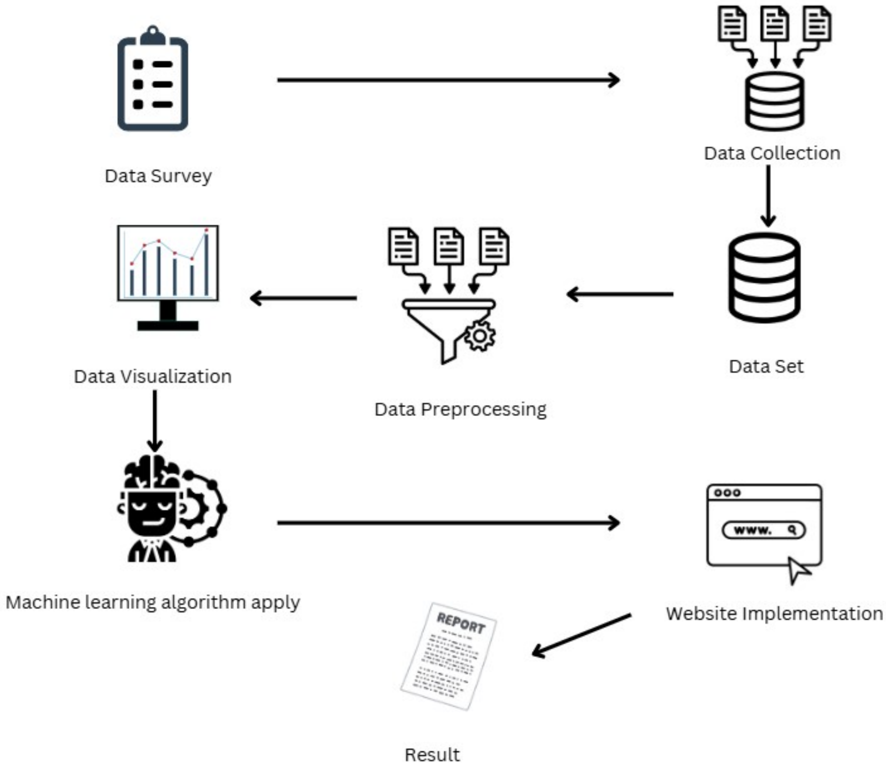


Figure 3.1: Proposed Methodology.

3.1.3 Functional and Nonfunctional Requirements

There is a set of Objectives that were created for achieving the goal, serving the purpose of making the drug addiction risk analysis website functional, non-functional, usable, efficient, and easily manageable. The functional requirements include giving the ability to enter demographic, psychological and environmental data at the user interface level. This data should then be fed through trained machine learning models and an easy to interpret addiction risk level output. Besides, it must allow users to analyze results either in the form of graphs or any other tools, so as to improve understanding. The system should permit updates to datasets and retraining of models for the benefit of the administrative users.

The non-functional requirements are aimed at achieving the website's purposes of being easily comprehensible, easily navigable and as secure as possible. Of course, it has to be easy to navigate, which is evident from the stripped down design of the website. It is essential for users to have access to the information anytime for any reason; thus, it is necessary to adapt the site for both PC and mobile-end and support multiple languages. One cannot compromise user identifiers and personal information, which means that the safety features such as data encryption and anonymization of information must be developed.

3.1.4 Context Diagram

The figure: 3.2 diagram also shows how the whole Drug Addiction Risk Analysis System interacts with some of the external entities. Here they are, Primary users: General users and administrators of the system. Owing to its simplicity, general users are able to insert personal and or demographic information into the system via the web based graphical user interface. This data is then analyzed by the system through applying some form of predictive modeling to evaluate the user as to the probability of being an addict and the system generates a report that the user can study. While, administrators are responsible for managing the system by modifying datasets, checking up on system performance, and managing the model characteristic. The diagram is used to illustrate these connections and further shows how the system mediates between data inputs and practical outputs that will make sense for the user and be easy to manage for the administrator.

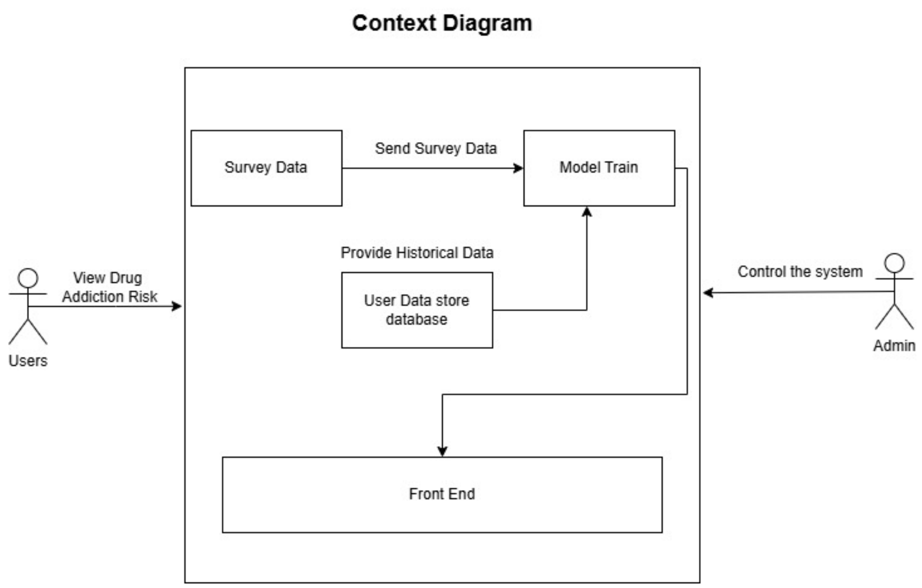


Figure 3.2: Context Diagram of Website.

3.1.5 Data Flow Diagram Level 1

The Data flow diagram 3.3 (DFD) shows the process of producing a Drug Addiction Risk Analysis System. The system involves interactions with two primary actors: User and Admin. Users supply raw input data and the Data Preprocessing component transforms this data into usable structures which are stored in the User Data Store for later use. The Machine Learning Model reads the historical data from the data store and with help of data cleaning mechanisms determines predicted risk levels. The results are then passed on to the Result Display component where it displays the risk analysis in view of the user. Admins use it by setting up the machine learning model and using the model to manage the data. Besides, they can also analyze raw results presented by the Result Display component and obtain general and detailed reports on these analyses.

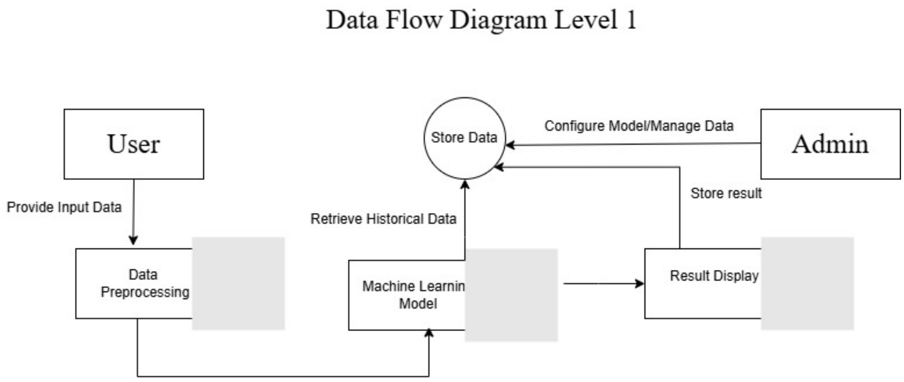


Figure 3.3: Data Flow Diagram Level 1 of Website.

3.1.6 UI Design

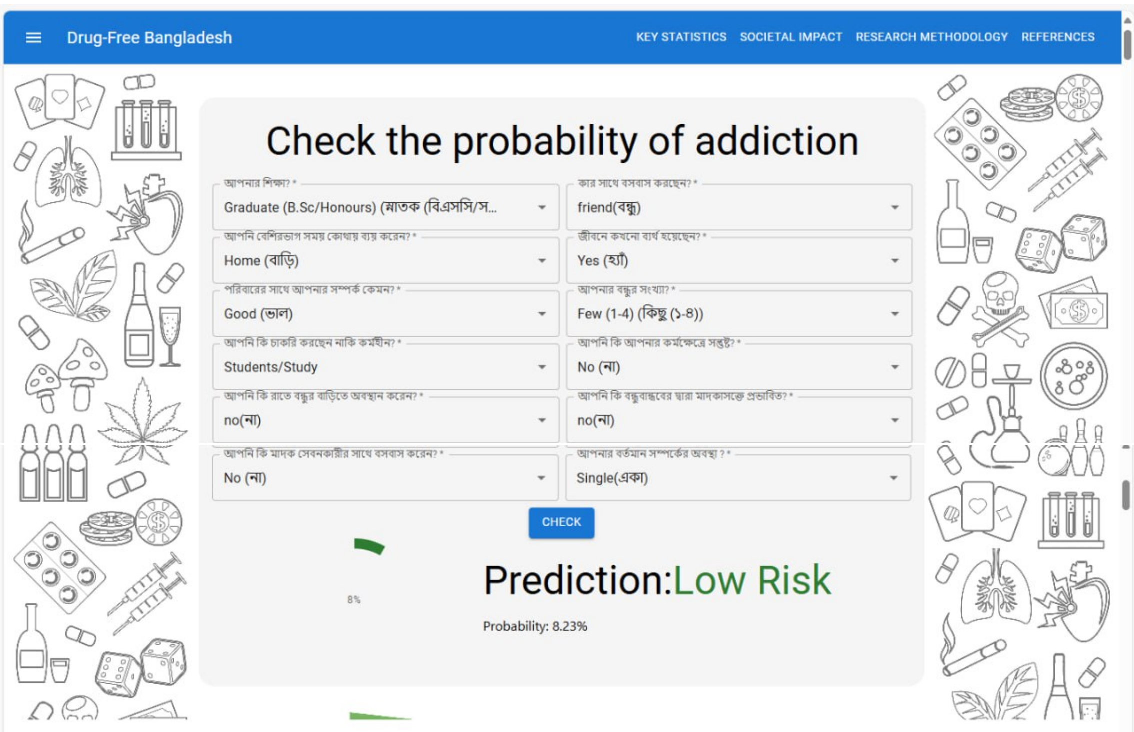


Figure 3.4: UI Design for user-friendly Website-1.

The project intends to offer a user-friendly web interface that allows users to analyze their drug addiction risk easily. The design of the website is built around simplicity and accessibility, guaranteeing that users can easily browse through the features and functionalities. Font-end design is particularly vital for a website where a user directly interact with the web site. Figure 3.4 show that their are a user interface where user give some question's answer like their education, friend numbers, where they can spend most of their time, is they are failure their life any time etc. total 12 questions. Based on this question our model predict the Drug addiction Risk Level and Produce the prediction result. Figure 3.5 section aims at giving an overview of our project including the mission,

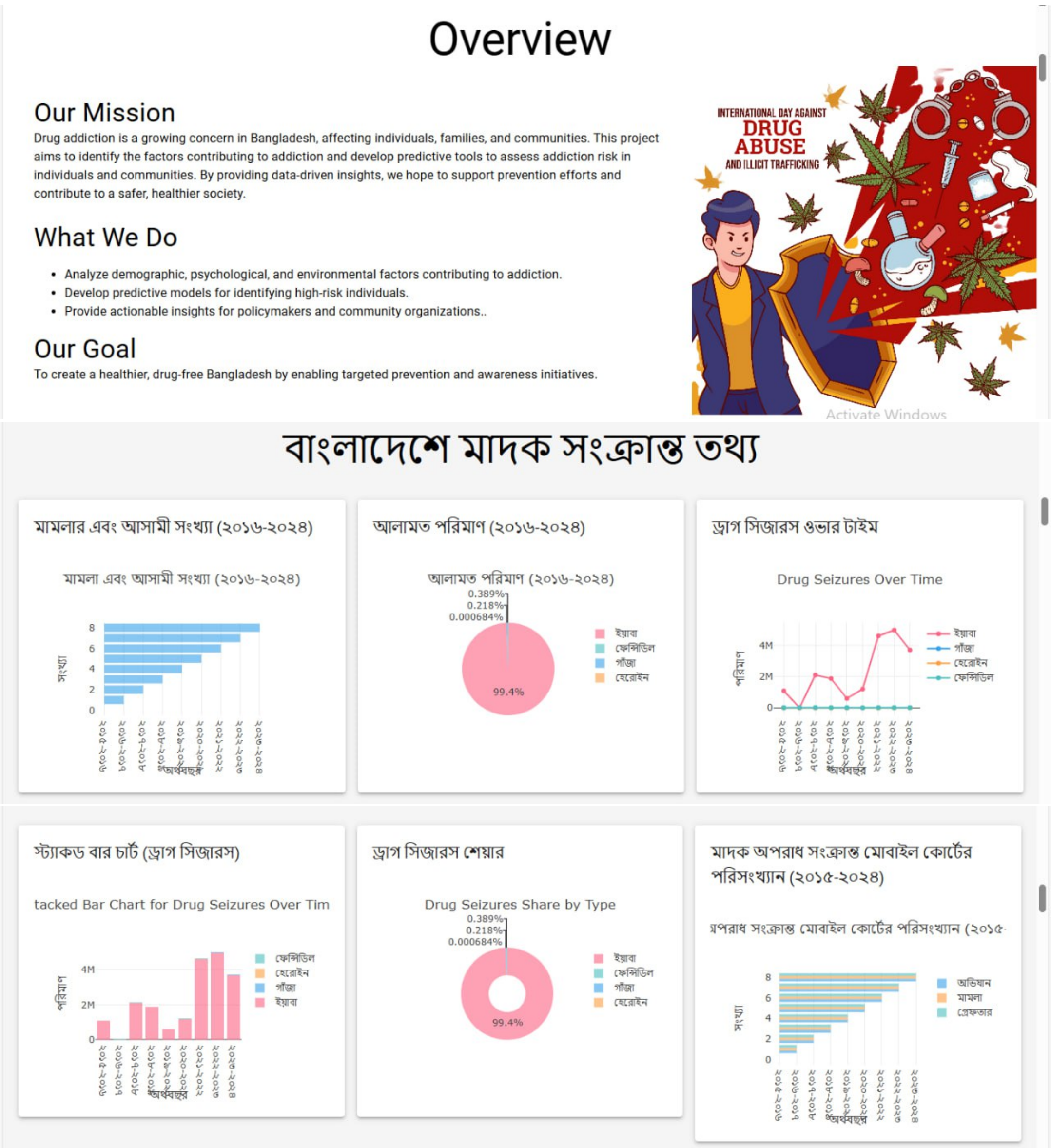


Figure 3.5: UI Design for user-friendly Website-2.

objective of the project as well as the overall project goals. Due to this, it gives a brief description through which one can be able to know the goal and aim as well as the general overview of the project. Also appended to this figure is essential information related to drugs in Bangladesh to help put the data into perspective for the consumers. Figure 3.6 importantinsights and present the learned important insights regarding our research study in an orderly manner. This allows the user to easily gain an understanding of the most important results of the study and also ensures convenience. In combination, these graphics help us as giving a comprehensive view of our project and its results for various types of users. These posts provide an efficient overview of the project's purpose and research outcomes by compiling the key pieces of information into concise overviews.

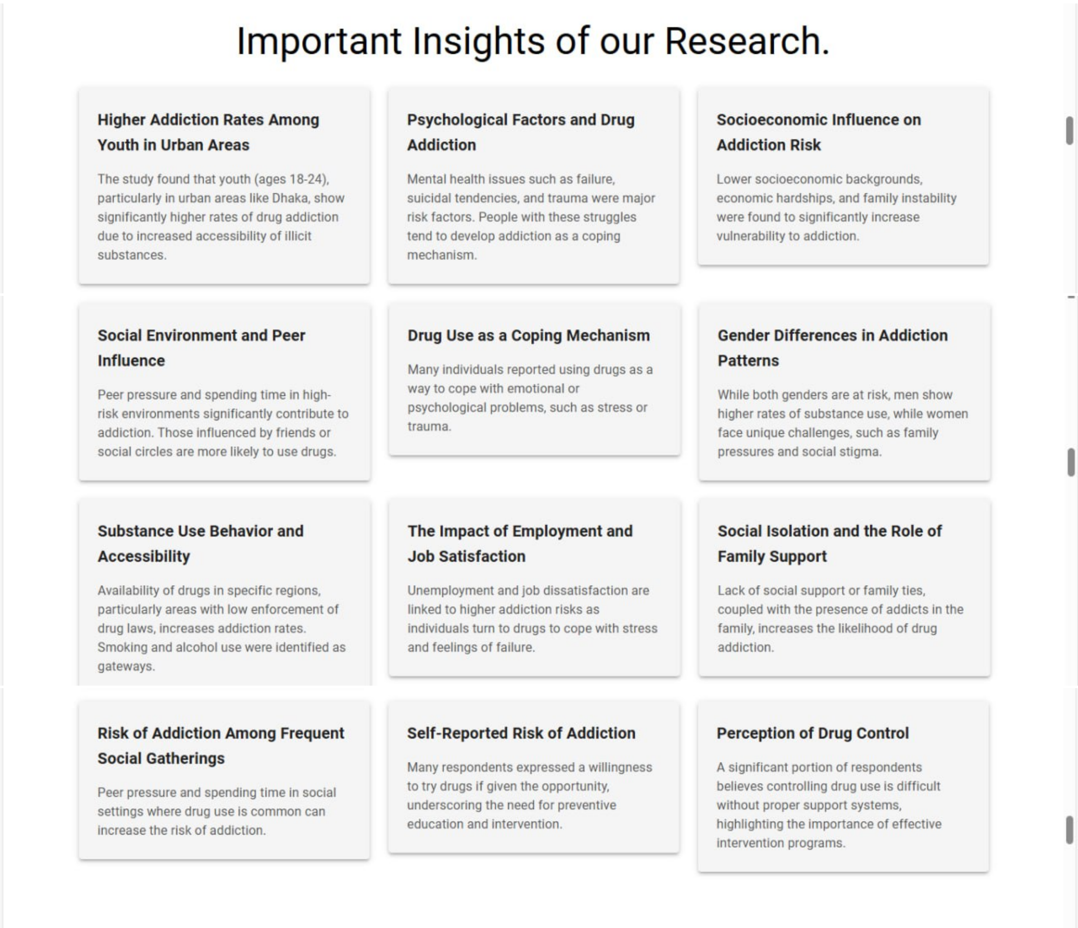


Figure 3.6: UI Design for user-friendly Website-3

3.2 Detailed Methodology and Design

Data Collection Procedure: Creating a dataset was one of the most challenging tasks in this study. Data was collected both online and offline survey. We collect 577 data using online survey through google form using questions like " Your age? Your gender? Your education? Living with? What is the purpose of your drug use? Where do you spend most of your time? Do you think you're a failure in life? Do you have mental or emotional problem? Do you have suicidal tendencies? Do you have social inhabitations or social pressure? Your family relationship? Your financial status of family? Is there drug addicts in your family? Number of your friends? Are you doing job or jobless? If you doing job then are you satisfied with your workplace? Do you have any litigation issues or case in court? Do you live with a drug addict? Do you smoke? Do you ever take drugs? What is your level of drug use? Your current relationship status? Do you stay at a friend's houses at night? Are you influenced by friends to become drug addicts? Would you take the drug if you got the chance? Do you think that it is possible to control the use of drugs? these questions. One Important things we can not take any personal information of our candidate like name, phone etc. so that they can give accurate data. Total 727 people participate our survey different age, different gender and different class. **Data Preprocessing:** First of all, we have to rename the features name because every feature name is so big that create problem when we work on code. Then we check null values and duplicate value and handle them, after that we have to convert the result into numeric value using mapping. This is our basic data preprocessing steps

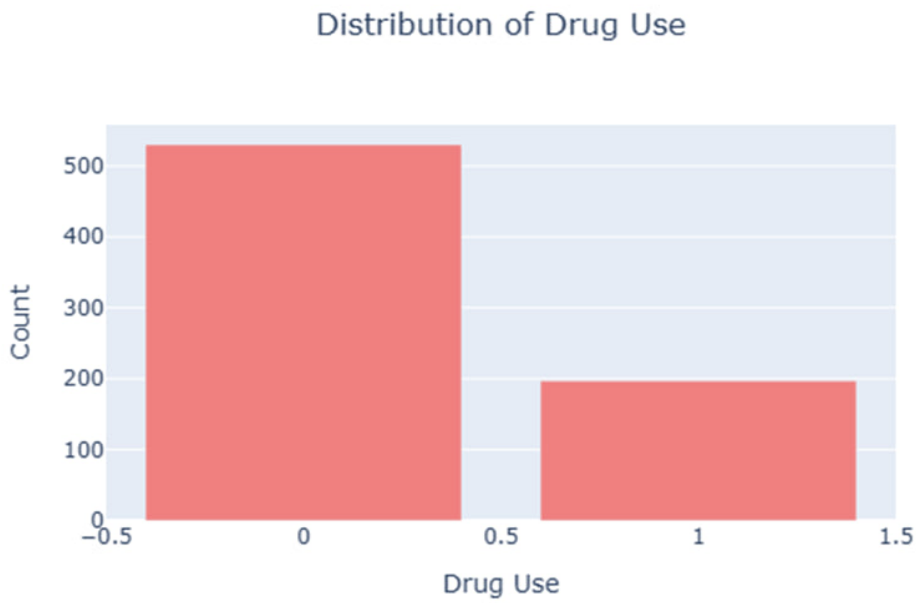


Figure 3.7: Distribution of Drug Use.

This figure 3.7 saw the distribution of drug use. There are two parts, one is counting data and another is drug use or not. We can see 517 people do not use drug and 200 people use drug. Feature Extraction: This step is very important for this research because we asked total 26 questions of every people among them important features we use for our model training process. We use two methods for Feature Engineering one is called correlation matrix and another is p-values from a chi-squared test.

The figure 3.8 correlation matrix is this. All Together, There Are 26 Variables. The cor-

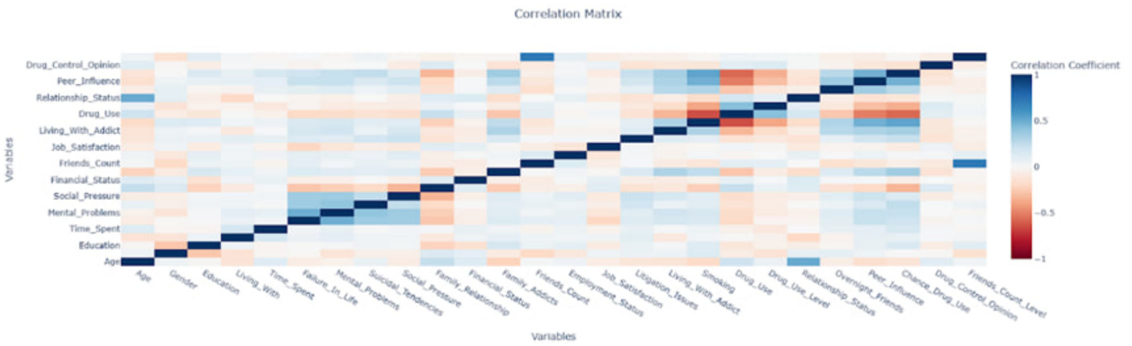


Figure 3.8: Correlation Matrix.

relation matrix shows relationships between many of the parameters and drug use, with color shade indicating the strength of the correlation, darker shades indicating stronger relationships, blue indicating positive relationships, and red negative. Main findings indicate a strong positive link between 'Drug Use' and 'Living with Addict,' and persons living with addicts are more prone to drug use. We find a moderate negative relationship between "Drug Use" and "Job Satisfaction," indicating that when the former rises, the latter decreases. The influence of peer pressure is positively associated only moderately with 'Drug Use'. A weak positive connection of 'Mental Problems' with 'Drug Use' suggests mental health problems are related to increased use of drugs. This is important direction for developing fitted strategies to prevent and reduce drug associated obstacles.

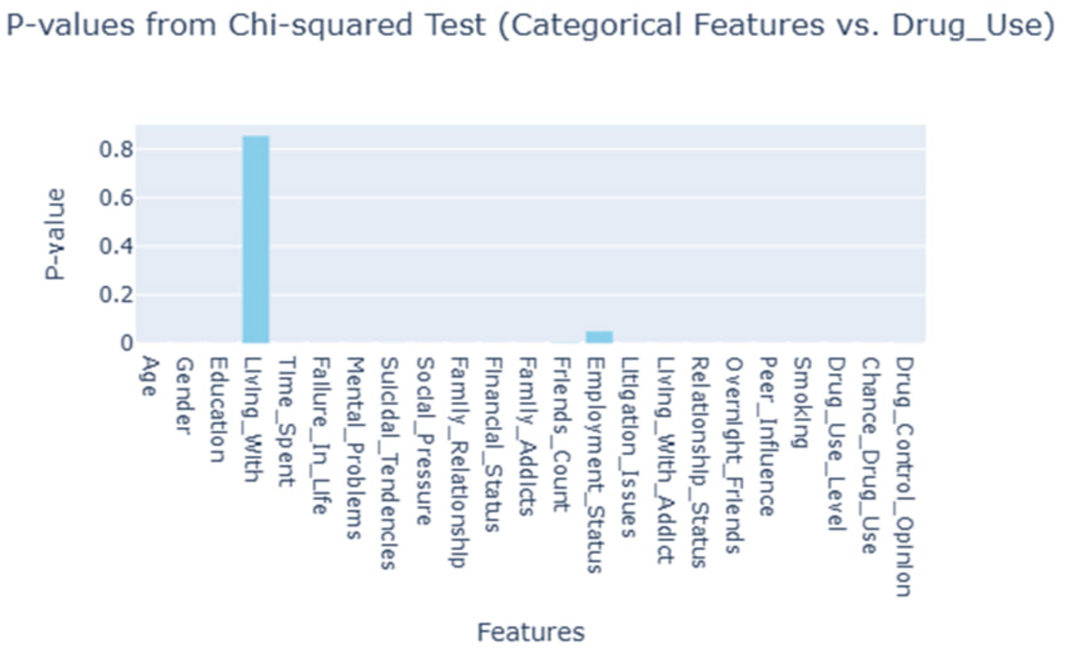


Figure 3.9: P-Values from A Chi-Squared Test.

The figure 3.9 displays p-values from a chi-squared test, evaluating the statistical significance of the association between category variables and drug usage. Features exhibiting p-values below 0.05 are deemed statistically significant, indicating a correlation with drug usage. The graphic indicates that only "Education" possesses a p-value below 0.05, signifying a statistically significant correlation with drug use. This shows that education level might be a factor influencing drug use. Dataset Split: The dataset was divided into training, validation, and testing sets. A common split ratio of 70% for training, 30% for testing was used. This section ensures that the model can be effectively trained and evaluated on invisible data to assess its generalizability.

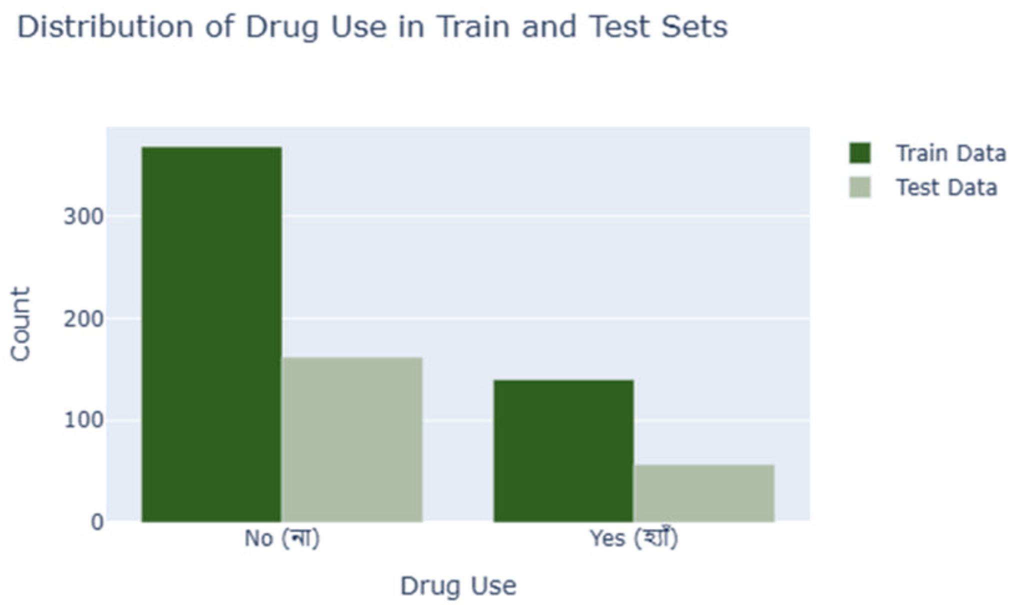


Figure 3.10: Dataset Splitting Ratio.

This figure 3.10 explains the train model total data is 508 and test model total data is 219.

Model Training: Various machine learning models were trained on the processed datasets. For traditional machine learning, algorithms such as LR, DT, RF, Naive Bayes, SVM were used.

Logistic Regression: Logistic Regression (LR) is a basic machine learning technique that is commonly applied for binary classification tasks, including drug addiction categorization in the context of our research. LR is governed by modeling the probability that a given input belongs to a particular class, commonly represented as a binary result. Mathematically, LR applies the logistic function (sigmoid) to a linear combination of input characteristics, which is stated as:

$$P(y = 1 \mid x; w) = \frac{1}{1 + e^{-(w^T x + b)}} \quad \text{(i)}$$

The LR model is a cost function, generally trained by minimizing logistic losses or cross-entropy losses, that punishes errors between projected probability and true labels in training data. Once trained, the LR model can make predictions by setting predictable probability thresholds, assigning positive class instances above a specific threshold and negative class instances below. Naïve Bayes: Naive Bayes is a probabilistic machine learning approach based on Bayes' theorem, which implies that attributes are conditionally independent due to class labels. Despite its simplicity, Naive Bayes generally performs well in classification problems, which makes it useful for drug addiction. Mathematically, the Naive Bayes classification predicts the probability of a class labeling y given a collection of qualities using Bayes' theorem:

$$P(y \mid x_1, x_2, \ldots, x_n) = \frac{P(y) \cdot P(x_1, x_2, \ldots, x_n \mid y)}{P(x_1, x_2, \ldots, x_n)} \quad \text{(ii)}$$

Naive Bayes simply assumes that all properties are given the conditionally independent class label y. Support Vector Machines: In the field of drug addiction, Support Vector Machine (SVM) stands out as a powerful classification method that is known for its ability in controlling the boundaries of high-dimensional data and nonlinear decisions. SVM works by translating the input data points to a high-dimensional feature space and determining the best hyperplane that splits the data points into various classes. This hyperplane is chosen to maximize the margin, which is the distance between the hyperplane and the nearest data point of each class, known as the support vector. Mathematically, the decision boundary of an SVM can be expressed as:

$$f(x) = w^T x + b \quad \text{(iii)}$$

In this study, SVM aims to classify drug addiction into positive or negative. The algorithm learns to find the optimal hyperplane that separates the feature space into distinct feeling classes.

Model Evaluation: The trained models were evaluated using standard metrics such as accuracy, precision, recall, and F1-score. In addition, the Confusion Matrix was developed to provide a detailed understanding of model performance across different sentiment classes. Accuracy: Accuracy is an important parameter for evaluating the performance of the prediction drug addiction model. It measures the ratio of correctly categorized examples to total instances and gives a straightforward approach to evaluate the overall efficacy of a model. The mathematical formula for accuracy is given by:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad \text{(iv)}$$

Precision: Precision is another key factor in evaluating the performance of the model. It quantifies the accuracy of positive (or negative) predictions provided by the model, offering insight into its ability to prevent erroneous positive (or negative) classifications.

The mathematical formula for accuracy is given by:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{(v)}$$

Recall: Recall is a basic parameter in evaluating the effectiveness of model. It assesses the model's capacity to recognize all relevant instances of positive (or negative) feelings, hence offering insight into its ability to prevent false negative classifications. The mathematical formula is given to remember:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{(vi)}$$

F1-score: The F-1 score is a significant parameter in evaluating the overall success of the model. It combines both accuracy and recall into a single rating, which provides a balanced evaluation of a model's correctness. F1 - The mathematical formula for the score is given:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{(vii)}$$

Model Selection: Based on the evaluation metrics, the best-performing model was selected. This model demonstrated the highest accuracy and robustness in high risk and low risk in the range of $0 - 100\%$.

Model Deployment: The final step is to deploy the selected model in a web application. Using Visual Studio code, the model was integrated into a express framework for the backend, while HTML, CSS, and JavaScript and React were used for the frontend design.

Statistical Analysis: In this study, statistical analysis plays an important role in verifying the effectiveness of various sentiment analysis models and understanding the underlying patterns within the dataset.

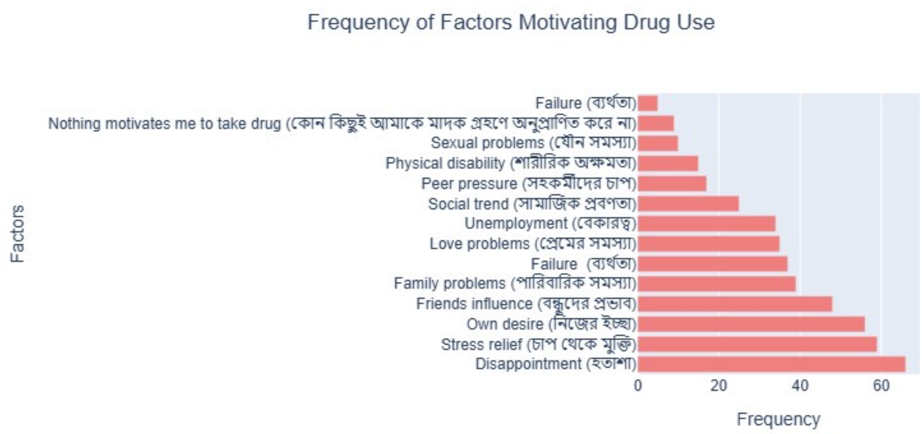


Figure 3.11: Motivating to Use Drugs Bar Chart

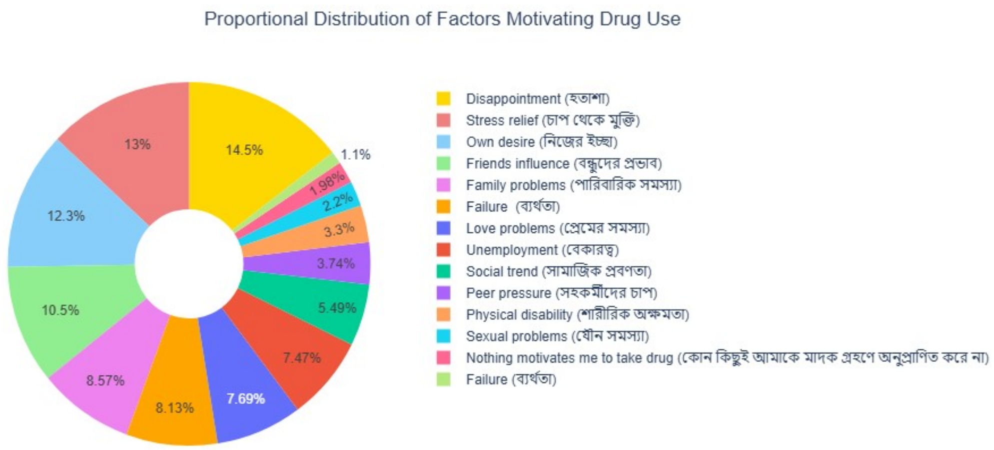
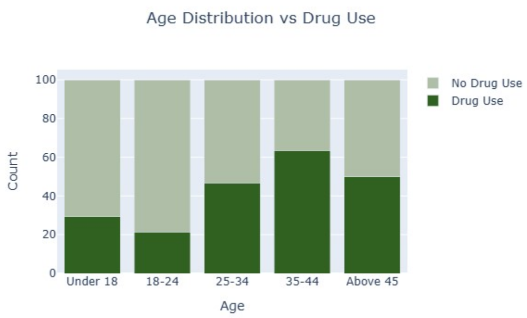


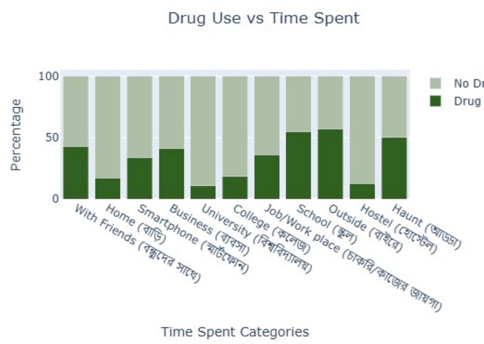
Figure 3.12: Distribution of Factors to Use Drugs Bar Chart

There are exactly fourteen factors for drug use. Maximum people use drugs for those factors. Even the disappointment factor is the highest rank of the motivating for drug use.So in this case we use a total of 10 models to find out the total number of drug factors rank. So those are here: This figure 3.11 and 3.12 describe the fourteen factors that are motivating to use drugs in figure Disappointment is 66,Stress relief is 59,Own desire is 56,Friends influence is 48,Family problemsis 39,Failure is 37,Love problems is 35,Unemployment is 34,Social trend is 25,Peer pressure is 17,Physical disability is 15,Sexual problems is 10,Nothing motivates me to take drug is 9,Failure is 5.

Figure 3.13 shows two graphs and describes the age distribution and time vs drug use.This



(a) Age Distribution

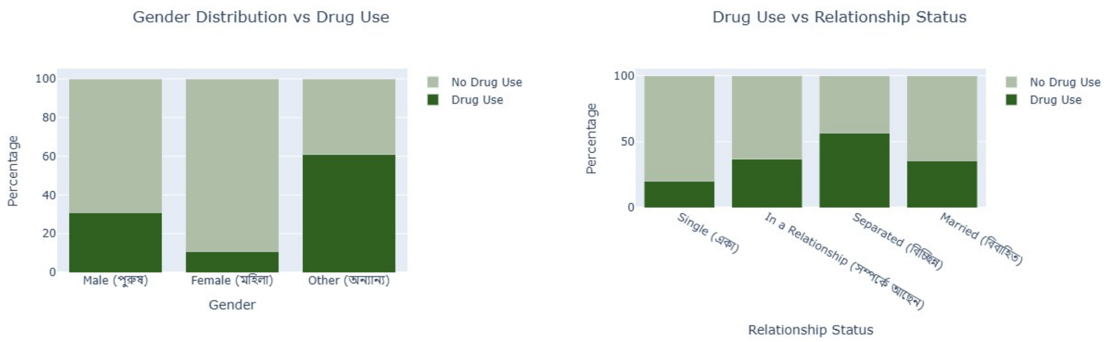


(b) Time Spent

Figure 3.13: Comparison of Age and Time Spent vs Drug Use

figure 3.13a shows age distribution vs drug use. There are two factors: Drug use and no drug use. An age starts from under 18 to above 45. The people of 29 % are addicted and the age is under 18. The bar rate for 35-44 age people is the highest to use drugs and the rating is 63%. Second highest 46% the ages are 25-34. We also see the age of 18-24 people are lowest rate of using drugs. The figure 3.13b shows "Drug Use vs Time Spent" and the percentage of drug use among different time-spent categories. The categories include activities like hanging out with friends, using a smartphone, being at home, and being in various locations like school, college, and workplaces. The graph reveals that drug use is higher in categories like "With Friends" and "Haun" compared to categories like "Home" and "Smartphone." This suggests that social settings and leisure activities might be associated with higher drug use. Even the highest number of people use drugs at school or outside and rate is 57 %.

Figure 3.14 shows two graphs and describes the Gender distribution and Relationship



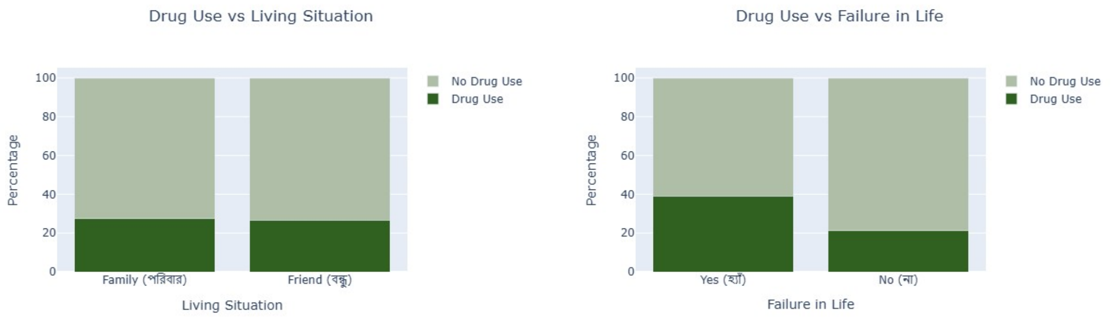
(a) Gender Distribution (b) Relationship status

Figure 3.14: Comparison of Gender Distribution and Relationship status vs Drug use

status vs drug use. This figure 3.14a shows three genders. Those are male, female and others. The highest number is 60%. These are the other gender people who use drugs. Second highest is 30%. These are the male who use drugs. Even we show that females are the lowest number that used drugs. The figure 3.14b mentions the Drug Use vs Relationship Status and shows the percentage of drug use among people with different relationship statuses. The categories on the x-axis are "Single," "In a Relationship," "Separated," and "Married." The graph shows that drug use is highest among people who are "In a Relationship" and "Separated." This suggests a potential correlation between certain relationship statuses and increased drug use.

Figure 3.15 shows two graphs and describes the comparison of Living Situation and Failure in Life vs Drug use. This figure 3.15a shows the Drug use vs Living situation. In this case their percentage is similar to drug use. About 27% of people use drugs while living with family or friends. The figure 3.15b about "Drug Use vs Failure in Life" and also shows the percentage of drug use among people who have experienced failure in life. The categories on the x-axis are "Yes" (indicating failure) and "No" (indicating no failure). The graph shows that drug use is significantly higher among individuals who have experienced failure in life. This suggests a potential correlation between failure and increased drug use.

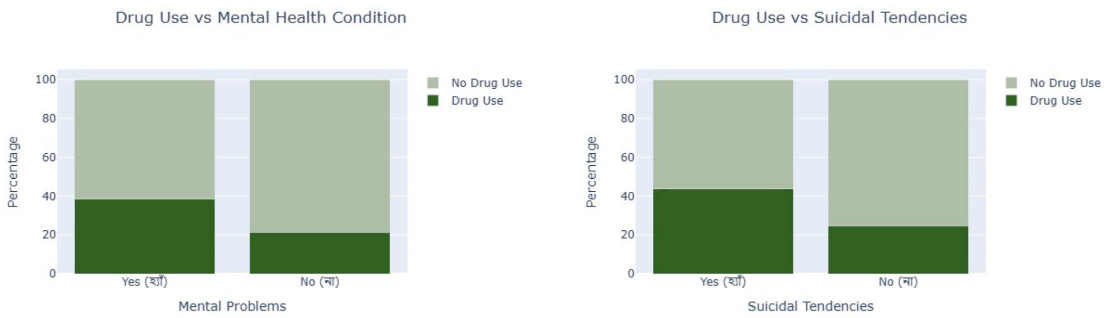
Figure 3.16 shows two graphs and describes the comparison of Mental Health Condition and Suicidal Tendencies vs Drug use. The figure 3.16a shows about Drug Use vs Mental



(a) Living Situation (b) Failure in Life

Figure 3.15: Comparison of Living Situation and Failure in Life vs Drug use

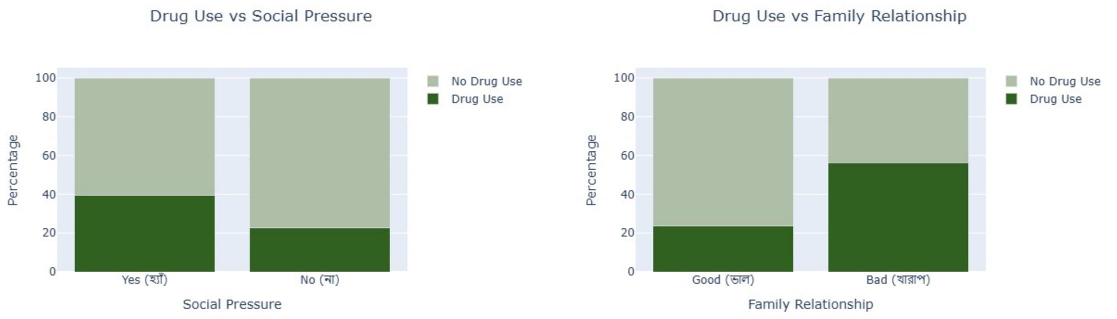
Health Condition and describe the percentage of drug use among people with and without mental health problems. The categories on the x-axis are "Yes" (indicating mental problems) and "No" (indicating no mental problems). The graph shows that drug use is significantly higher among individuals with mental health problems. This suggests a potential correlation between mental health issues and increased drug use. The figure 3.16b cribes the Drug Use vs Social Pressure and also shows the percentage of drug use among people who experience social pressure and those who don't. The categories on the x-axis are "Yes" (indicating social pressure) and "No" (indicating no social pressure). The graph shows that drug use is higher among individuals who experience social pressure.



(a) Mental Health Condition (b) Suicidal Tendencies

Figure 3.16: Comparison of Mental Health Condition and Suicidal Tendencies vs Drug use

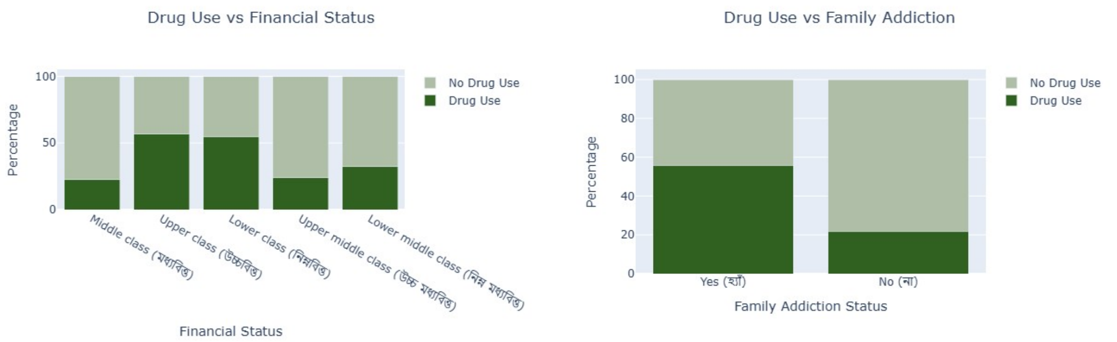
Figure 3.17 shows two graphs and describes the comparison of Social Pressure and Family Relationship vs Drug use. The figure 3.17a cribes the Drug Use vs Social Pressure and also shows the percentage of drug use among people who experience social pressure and those who don't. The categories on the x-axis are "Yes" (indicating social pressure) and "No" (indicating no social pressure). The graph shows that drug use is higher among individuals who experience social pressure. The figure 3.17b mentions the Drug Use vs Family Relationship and also mentions the percentage of drug use among people with



(a) Social Pressure (b) Family Relationship

Figure 3.17: Comparison of Social Pressure and Family Relationship vs Drug use

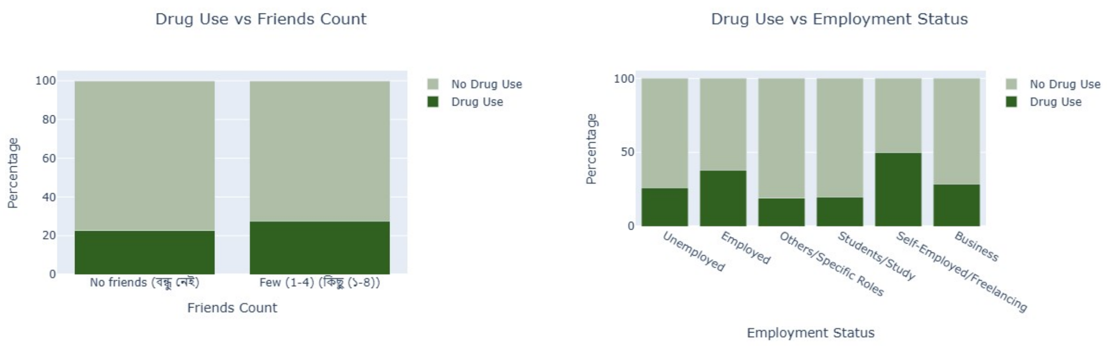
good and bad family relationships. The categories on the x-axis are "Good" and "Bad." The graph shows that drug use is significantly higher among individuals with bad family relationships. This suggests a potential correlation between poor family relationships and increased drug use.



(a) Financial Status (b) Family Addiction

Figure 3.18: Comparison of Financial Status and Family Addiction vs Drug use

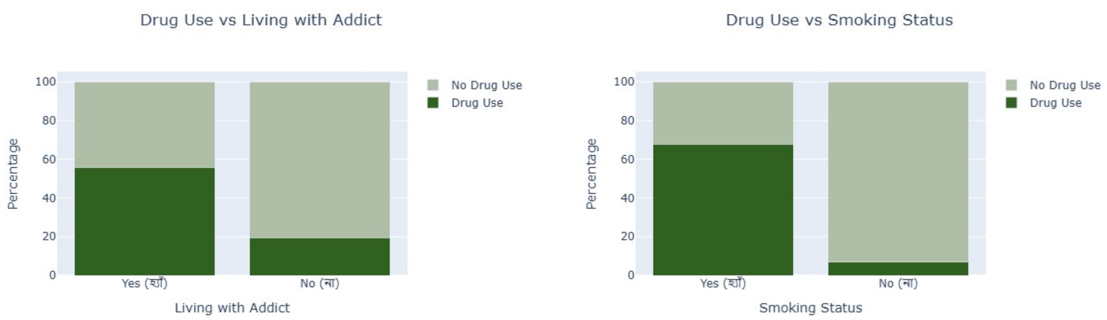
Figure 3.18 shows two graphs and describes the comparison of Financial Status and Family Addiction vs Drug use. The figure 3.18a shows the Drug Use vs Financial Status also shows the percentage of drug use among people from different financial classes. The categories on the x-axis are "Middle Class," "Upper Middle Class," "Lower Middle Class," "Upper Class," and "Lower Class." The graph shows that drug use is highest among the "Lower Middle Class" and "Upper Middle Class" individuals. The figure 3.18b shows the Drug Use vs Family Addiction and also shows the percentage of drug use among people with and without family members who have addiction issues. The categories on the x-axis are "Yes" (indicating family addiction) and "No" (indicating no family addiction). The graph shows that drug use is significantly higher among individuals with family members who have addiction issues. Figure 3.19 shows two graphs and describes the comparison of Friends Count and Employment Status vs Drug use. The figure 3.19a shows the Drug Use



(a) Friends Count (b) Employment Status

Figure 3.19: Comparison of Friends Count and Employment Status vs Drug use

vs Friends Count and also shows the percentage of drug use among people with different numbers of friends. The categories on the x-axis are "No friends" and "Few (1-4)." The graph shows that drug use is higher among people with few friends compared to those with no friends. This suggests a potential correlation between having a few friends and increased drug use. The figure 3.19b shows the "Drug Use vs Employment Status" and also shows the percentage of drug use among people with different employment statuses. The categories on the x-axis are "Unemployed," "Employed," "Others/Specific Roles," "Students/Study," "Self-Employed/Freelancing," and "Business." The graph shows that drug use is highest among "Self-Employed/Freelancing" individuals.



(a) Living with Addict (b) Smoking Status

Figure 3.20: Comparison of Living with Addict and Smoking Status vs Drug use

Figure 3.20 shows two graphs and describes the comparison of Living with Addict and Smoking Status vs Drug use. The figure 3.20a describes the Drug Use vs Living with Addict and also discusses the percentage of drug use among people who live with and without addicts. The categories on the x-axis are "Yes" (indicating living with an addict) and "No" (indicating not living with an addict). The graph shows that drug use is significantly higher among individuals who live with addicts. The figure 3.20b discusses the Drug Use vs Smoking Status and shows the percentage of drug use among people who smoke and those who don't. The categories on the x-axis are "Yes" (indicating smoking) and "No" (indicating not smoking). The graph shows that drug use is significantly higher among people who smoke compared to those who don't. This suggests a potential correlation between smoking and increased drug use.

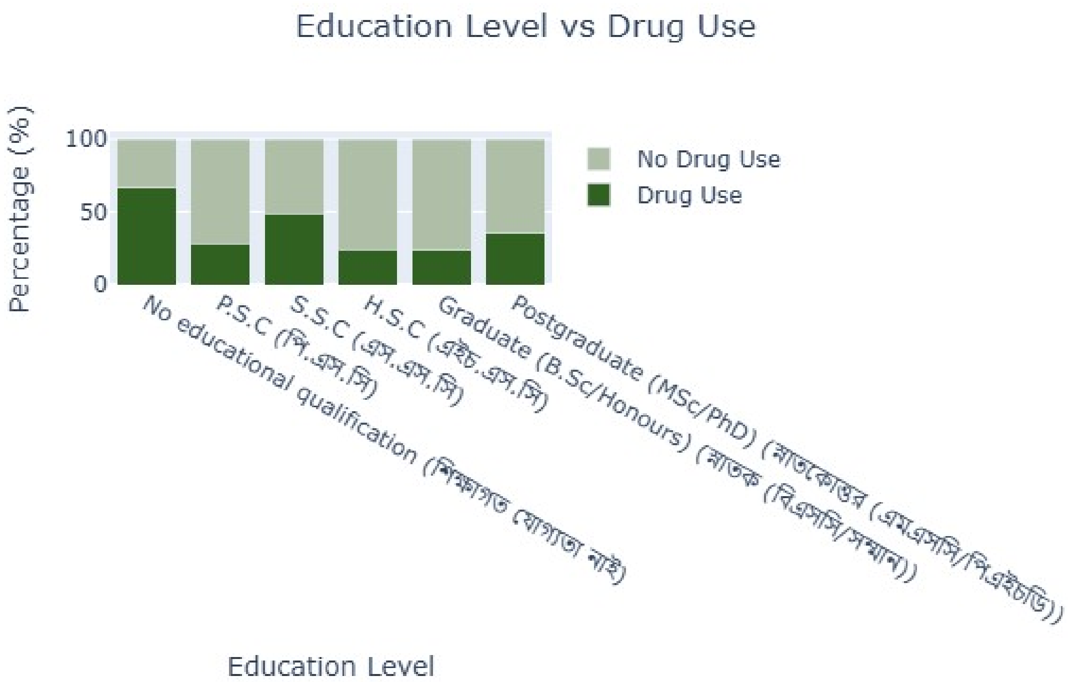


Figure 3.21: Education Level vs drug use Bar Chart.

This figure 3.21 shows education level vs drug use. There are 6 categories. the no educational people are the increasing number of uses drugs and their percentage is 66%. The equal levels are HSC and graduated candidate peoples and their percentage is 23%. Second highest number is SSC candidates. These people use drugs badly and the rating is 48%.

3.3 Project Plan

This project plan 3.22 is logically structured to ensure that the study will be completed on time and within the princely financial constraints. The first one is the project planning and scope definition, where the objective of the project are set, the timelines set and the resources required are identified. Structuring of questionnaire follows, which involves developing or conducting surveys that are well designed, reviewed to give accurate data. Online and offline surveys are given to potential respondents and responses collected are part of data collection. Data pretreatment subsequently focuses on the structuring, clearing and preparing the dataset for analysis. During the model creation, various machine learning techniques are applied and used so as to arrive at the right models. Performance measures are then evaluated. This is an assurance of conclusions after model interpretation as well as making the conclusion End users to understand. As such, the project also includes development and deployment of an easy to navigate website in addition to the integration of the predictive models for practical purposes. Finally, reporting and documenting also include preparing detailed reports, doing presentations of the results accomplished and offering realistic suggestions. This well developed plan ensures compliance with the objectives of a project while offering the foundation for an effective, research based method in handling the risks involved in substance dependency.

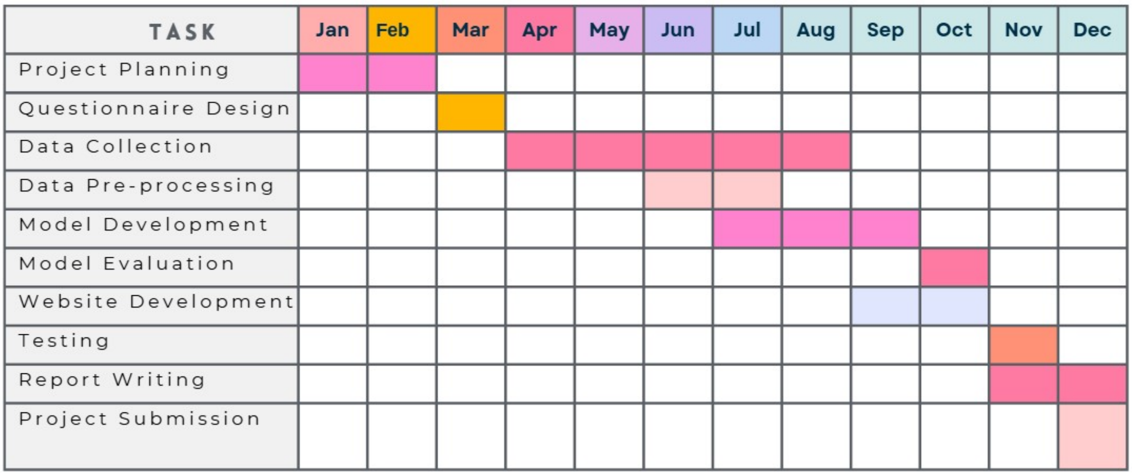


Figure 3.22: Project Plan Grantt Chart

3.4 Task Allocation

For the purpose of efficiency and use of knowledge, the specific tasks of this research project are divided and assigned amongst the different members of the team. The project manager oversees the progression of the project, is held accountable for the completion of project specific deadlines and is responsible for project planning and the definition of project scope. For the purpose of validity and generalisability of the questions, the survey methodology is developed by the experts of the area. Surveys are completed by field workers together with the help of survey coordinators who are also responsible for collecting answers as a part of the data gathering process. The task of data cleaning and collection, organizing data, and preparing data for analysis is performed under the Supervision of data scientists. Also, they create and compare models, select and fine-tune the machine learning techniques to get the best outcome. Designing and developing a user-friendly website, including both visually appealing front-end design work and back-end model addition is the job of web developers. Communication specialists assist in documentation and reporting because they prepare documents and demonstrations containing summary of outcomes and potential messages. That is why this clear role allocation is each team member is assured of concentration on their area of specialization; which enhances co-operations with goals for timely project completion.

3.5 Summary

The specific context of this study lies in the development of an affordable and large scale drug addiction risk analysis system for the metropolitan region of Bangladesh with Python and machine learning. The study progresses through three phases: gather data through surveys, use the model assessment step, and place the best performing model on a website platform. Handling of null and duplicate values is the initial step of data preprocessing that is proceeded by feature engineering. Scheme Using & Training/Evaluation Metrics & Algorithms LR & DT & RF & Naive Bayes & Numeric Complete KNN & SVM & SGD & Numeric & Complete HTML, CSS, React, and Express.js programming make up the website that contains direct interfaces to the implement predictive models as well as provide secure usage for the users. The system also enables users to estimate dependency probability regarding demographic, psychological data, and environment; the updates and performance are controlled by administrators. The 52 week long project costs BDT 1,37,000 to BDT 1,61,000 and includes, sound planning strategy, division of tasks based on the participants' roles and responsibilities and advocacy for data anonymity and encryption. This is the basis of a sound approach for managing substance dependency risks.

Chapter 4

Implementation and Results

4.1 Environment Setup

For this study, we conducted it in a google colab environment, and used python as our main programming language. The study is divided into three parts. Data had not yet been collected. We got the data using Google forms and manual physical surveys. Data analysis, model selection and model evaluation were the second stage. This project requires little computing power, and with the help of Google colab we get provided with the necessary computing infrastructure, such as a Tesla GPU and up to 16 GB of RAM, which is enough. For data analyzing we used Panda Python package which is a NumPy package. Furthermore, I implemented machine learning methods in Skeetlearn library with the use of the Logistic Regression (Baseline Model) as the first model, Decision Tree and Random Forest, Support Vector Machine (SVM) Neve Base, AdaBoost (Adaptive Boosting), Nearest Neighbors (KNN), MLPClassifier. In the DNA of part of the study placing top performing models on a website were built using the Visual Studio Code environment.

4.2 Comparative Analysis

The table 4.1 provides a side by side analysis of the machine learning approaches employed by different researchers with regards to the plans for developing solutions for predictive modeling. Classic machine learning algorithms such as LR, DT, and RFs were evaluated by Rahman, et al.delivering the maximum of 79.6% accuracy. In the same year, Khan et al., used Multivariate Logistic Regression (MLR) and Support Vector Machines (SVM) classifiers and the MLR yielded an accuracy of 80%. Zhao et al. used Decision Trees and Random Forests, the authors managed to achieve a high accuracy of 78 percent using DT. Ghitza et al. on the other hand performed a comparative study of Logistic Regression, Decision Trees, Random Forests and Support Vector Machines; among them, SVM had the best mean accuracy of 81 percent. Nguyen et al. also employed CNN, LSTM, and Bi-LSTM models where LSTM model achieved the highest accurate of 87%. Similarly, Kumar et al. achieved 81% efficiency using a number of machine learning environments. Similar to authors interested in interpretable ML models, Xie et al. achieved the 87% accuracy as well. Also, in my work all kinds of algorithms were employed such as Logistic Regression, Decision Trees, Random Forests, SVM, Naive Bayes (NB), AdaBoost, KNN, MLP Classifier, Gradient Boosting, and CatBoost. Among them, AdaBoost and CatBoost achieved large accuracies of 0.8127; thus, the current experiments demonstrated reasonable performance compared to prior tests.

Table 4.1: Comparative Analysis My Work with Other Study

Study	**Method & Techniques**	**Results**
Rahman et al.	Logistic Regression, Decision Trees, Random Forests	LR = 79.6%
Khan et al.	Multivariate Logistic Regression, Support Vector Machines	MLR = 80%
Zhao et al.	Decision Trees, Random Forests	DT = 78%
Ghitza et al.	Logistic Regression, Decision Trees, Random Forests, Support Vector Machines	SVM = 81%
Nguyen et al.	CNN, LSTM, Bi-LSTM	LSTM = 87%
Kumar et al.	Machine Learning Frameworks	LR = 81%
Xie et al.	Interpretable Machine Learning Models	87%
My Study	LR, DT, RF, SVM, NB, AdaBoost, KNN, MLP Classifier, Gradient Boosting, CatBoost.	AdaBoost = 81.27% CatBoost = 81.27%

4.3 Results and Discussion

In this work, I use the experimental setting for this work to consist of a methodical combination of hardware, software, and data, all necessary for a thorough investigation of sentiment analysis. Accelerated processing workloads require high performance computing resource such as a powerful central processor unit (CPU) complemented by possibly a powerful graphics processing unit (GPU). Deep learning frameworks like TensorFlow, machine learning methods libraries (e.g., scikit-learn) in a Python environment are included in software requirements. Physical survey and online survey are two methods of the experimental dataset. Several different machine learning models such as logistic regression, decision trees, naïve bias, k nearest neighbors, support vector machines, etc. are used on this processed information and are trained on the same. The effectiveness of the model is completely evaluated with respect to accuracy, precision, recall, F-1 score and explanatory score as evaluation criteria. From the vetting process, the process is scrupulously ethical in considering ethical considerations such as data protection and compliance.

The figure 4.1 titled "Model Accuracy Comparison" shows the accuracy of various ma-

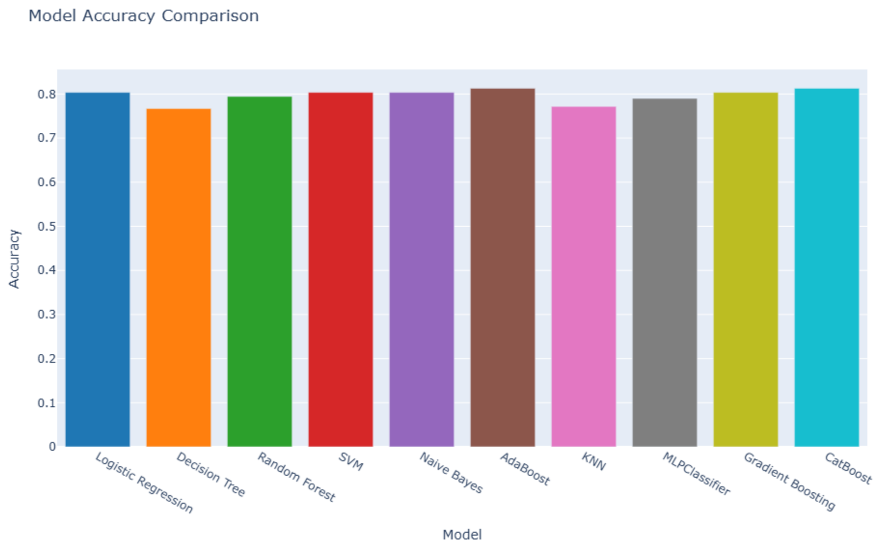


Figure 4.1: Model Accuracy Comparison

chine learning models. The x-axis lists different models like Logistic Regression, Decision Tree, Random Forest, and so on. The y-axis represents the accuracy score, ranging from 0 to 0.8. The graph visually compares the performance of these models, with CatBoost and Gradient Boosting appearing to have the highest accuracy scores.

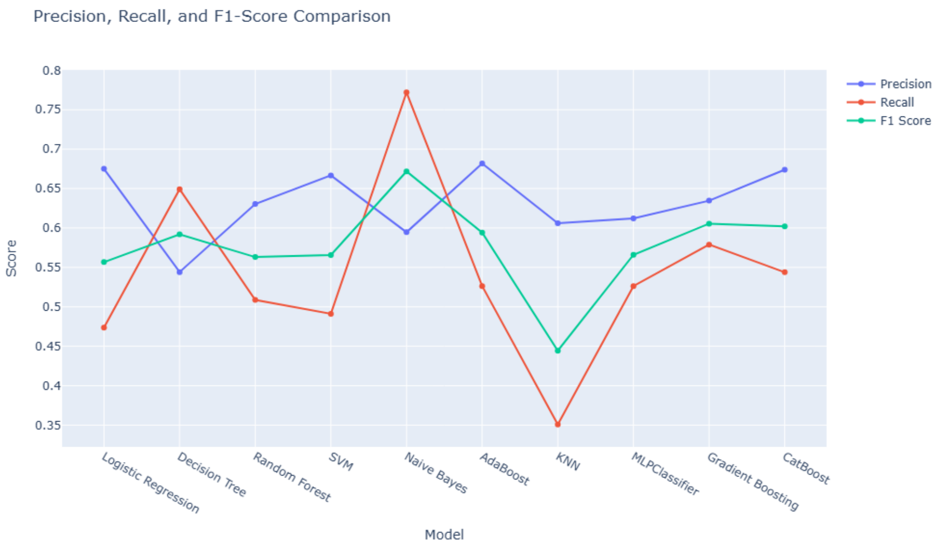
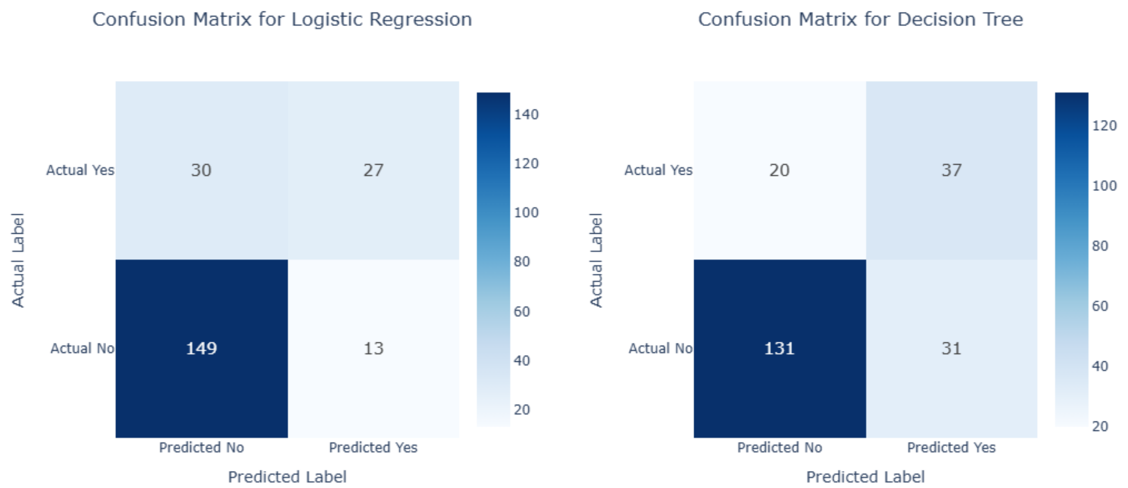


Figure 4.2: Precision, Recall, and F1-Score Comparison Graph

The line graph figure 4.2 displays the comparison of Precision, Recall, and F1-Score across ten machine learning models: It includes Logistic Regression, Decision Tree, Random Forest, SVM, Naive Bayes, AdaBoost, and KNN trained with MLPClassifier, Gradient Boosting Classifier and CatBoost. These models are shown on x axis and we have scores ranging from 0.35 to 0.8 on y axis. For KNN, the blue line, Precision, bobs between just under 0.55 (the low end) and 0.73 (the high end) for CatBoost. For Naive Bayes, recall peaks at 0.75 and drops very quickly to 0.35 for AdaBoost. We can see in the green line (F1-Score) that Gradient Boosting and Camping have around 0.65 while AdaBottom decreases to 0.45. The Recall shows significant fluctuation, with Naive Bayes' best score, though the F1-Score does not fully reflect this when this fluctuation is due to imbalance with Precision. Overall, Gradient Boosting and CatBoost are highly stable and balanced metrics and are strong performancs. Illustrating tradeoffs and model strengths when considering classification performance, this graph creates from data how the different models perform.

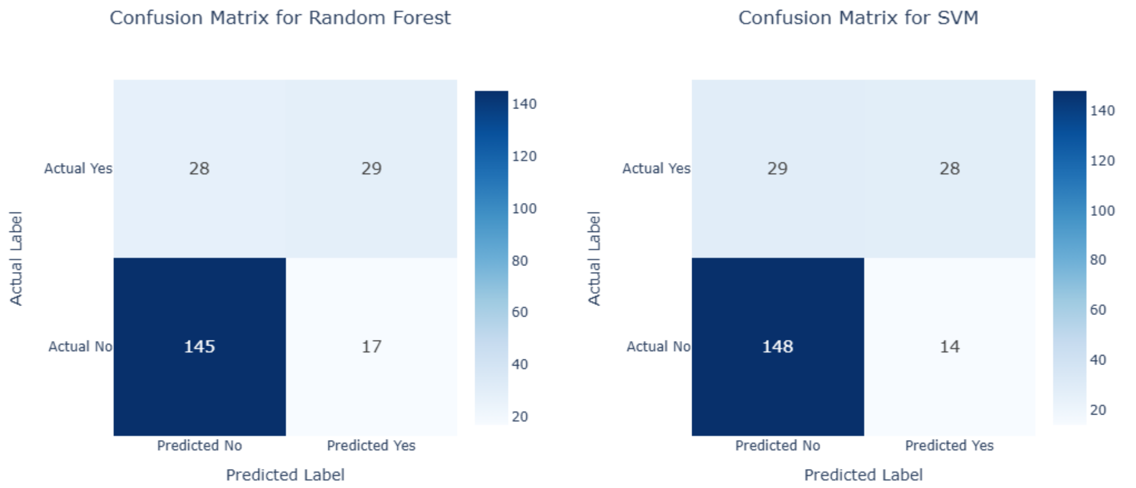
Figure 4.3a shows that the Logistic Regression (LR) model accurately predicts 174 data points but misclassifies 43. On the other hand, Figure 4.3b demonstrates that the Decision Tree (DT) model correctly predicts 168 data points, but it makes incorrect predictions for 51 sentences. These results highlight the varying performance of both models, with LR slightly outperforming DT in terms of accuracy, while DT has a higher misclassification rate. Such comparisons are crucial in determining the most suitable model for the task.



(a) Confusion Matrix for LR Algorithm (b) Confusion Matrix for DT Algorithm

Figure 4.3: Confusion Matrix for for LR and Dt Algorithm

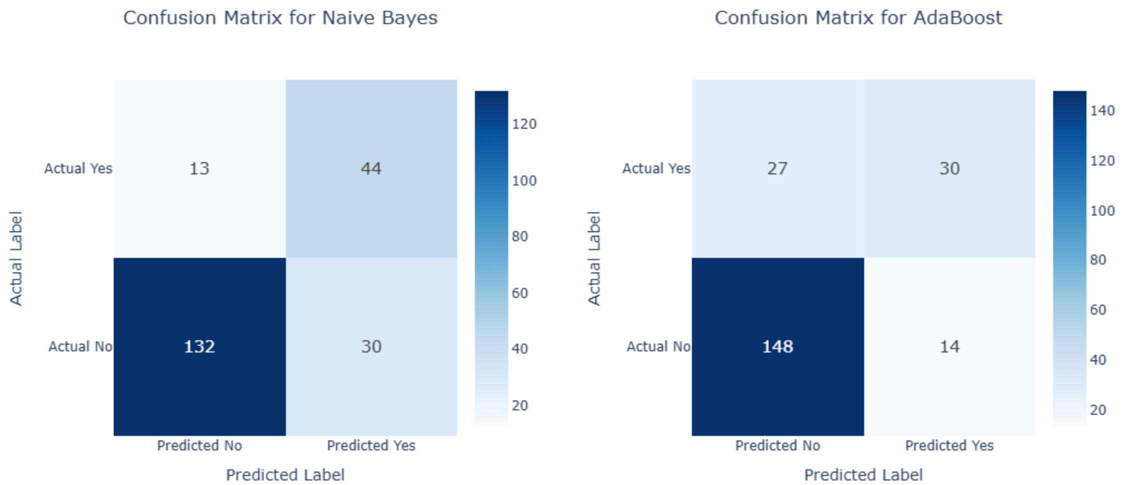
Figure 4.4a illustrates that the Random Forest (RF) model correctly predicts 174 data points, while it mispredicts 45 data points. On the other hand, Figure 4.4b demonstrates the performance of the Support Vector Machine (SVM) model, which accurately predicts 156 data points and makes incorrect predictions for 43 data points. Both figures highlight the performance of the respective models in terms of correct and incorrect predictions, providing insights into their prediction accuracy and error rates.



(a) Confusion Matrix for RF Algorithm (b) Confusion Matrix for SVM Algorithm

Figure 4.4: Confusion Matrix for RF and SVM Algorithm

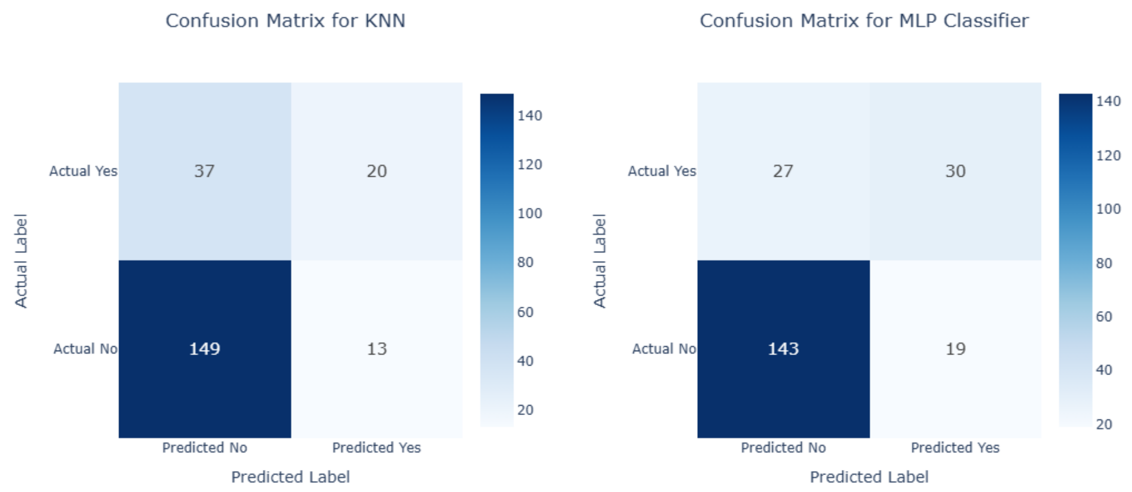
Figure 4.5a illustrates that the Naive Bayes (NB) model accurately predicts 176 data points, while misclassifying 43 data points. In contrast, Figure 4.5b demonstrates that the AdaBoost model correctly predicts 178 data points but makes errors in 41 data points. Both models perform well, but AdaBoost shows slightly higher accuracy than Naive Bayes. These comparisons highlight the effectiveness of each model in making predictions, with AdaBoost demonstrating a marginally better performance in terms of accurate classifications.



(a) Confusion Matrix for NB Algorithm (b) Confusion Matrix for AdaBoost Algorithm

Figure 4.5: Confusion Matrix for NB and AdaBoost Algorithm

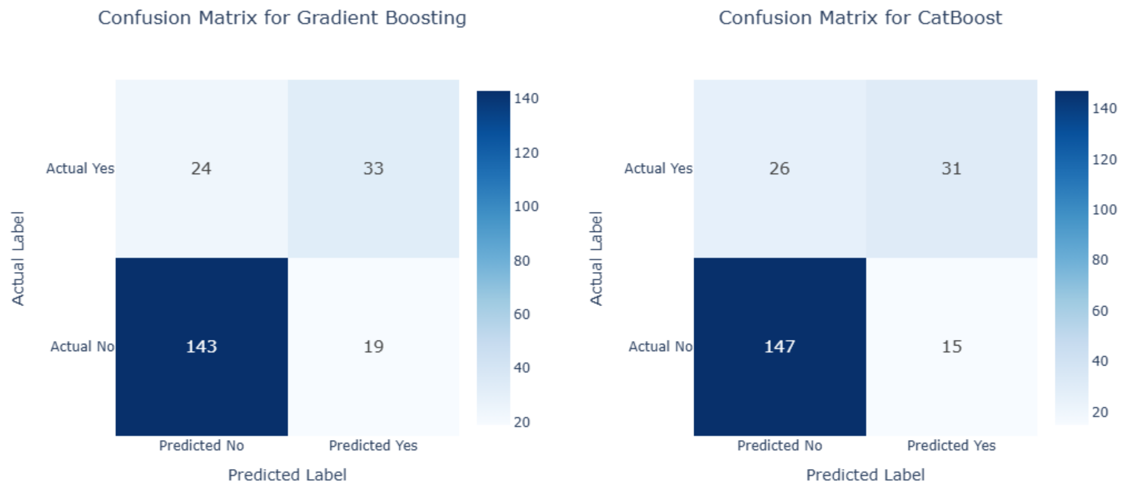
Figure 4.6a illustrates that the KNN model accurately predicts 169 data points but makes incorrect predictions for 50. In comparison, Figure 4.6b shows that the MLP model performs slightly better, accurately predicting 173 sentences, while misclassifying 46. These results highlight the predictive accuracy of both models, with the KNN model showing a lower accuracy compared to MLP. Despite the differences, both models demonstrate strengths and weaknesses in their performance, with MLP being slightly more accurate in this particular dataset.



(a) Confusion Matrix for KNN Algorithm (b) Confusion Matrix for MLP Algorithm

Figure 4.6: Confusion Matrix for KNN and MLP Algorithm

Figure 4.7a illustrates that Gradient Boosting correctly predicts 176 data points, but makes incorrect predictions for 43 data points. In contrast, Figure 4.7b shows that CatBoost achieves higher accuracy, correctly predicting 178 data points, with only 41 data points being predicted incorrectly. This comparison highlights the slight difference in performance between the two models, with CatBoost demonstrating a marginally better predictive accuracy over Gradient Boosting in this particular dataset. Both models show strong performance but have room for improvement in handling specific data points. Finally, we conclude that the accuracy and prediction performance of different machine learning models, displayed in the figures, differ. Figure 4.7a shows 176 correctly predicted data points with 43 wrongly predicted, while Figure 4.7b shows CatBoost predicts 178 data points correctly and makes 41 wrong predictions. It turns out that both models perform very well, and CatBoost marginally better works on this dataset. These findings underline the need of selecting a model based on the data and performance metric characteristics.



(a) Confusion Matrix for Gradient Boosting Algorithm

(b) Confusion Matrix for CatBoost Algorithm

Figure 4.7: Confusion Matrix for Gradient Boosting and CatBoost Algorithm

4.4 Summary

The review of machine learning models' performance on a given dataset was provided including an insight into the comparative accuracy of each. Using the existing computer resources and widely available software libraries, we successfully train and analyze models such as Scikit Learn. Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM) Naive Bayes, AdaBoost, K-Nearest Neighbors (KNN) Multi-Layer Perceptron (MLPC Classifier) Gradient Boosting, and CatBoost were combined with multiple machine learning methods. The two major parameter of evaluating accuracy were highest respectively about the values of 0.77 and 0.81. According to the test results, AdaBoost and CatBoost generated an accuracy score of 0.81 whereas Decision tree and KNN resulted in accuracy score of 0.77. Overall, we build models like Logistic Regression, SVM, Naive Bayes, and Gradient Boosting and we found out that they all came up with the same competitive accuracy score of 0.80, which shows they are trustworthy. Accuracy obtained for both Random Forest and MLPC Classifier with an accuracy of 0.79. Section 5.2 completes an extensive assessment of these results with special emphasis on the competitiveness of AdaBoost and CatBoost results. The models are compared to underline the merits and shortcomings of the models and the outlook for improvement.

Chapter 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

5.1.1 Software Standards

The standards used in this research can ensure that the web-based system developed to assess the risk of drug addiction is dependant, efficient and user-friendly. Coding standards are observed to an agreed set of practices ranging from the use of PEP 8 for coding standards for readability, maintainability and high standards in python source. Version control is managed by Git and GitHub, which allow the authors of the report and other interested parties track changes during development. The website also complies with the WCAG 2.1 accessibility standards sought to ensure the website is usable for people with additional needs. HTTPS deployment and data encryption are measures applied to security requirements to protect users data and maintain user secrecy. The software design approach used maintains modularity for scalability in the future for additions or connections to the design. Additionally, it is all the machine algorithms and statistical models respect the ethical AI principles which emphasize explaining and being fairly when making decisions. The deployment method also utilizes CI/CD pipelines to make upgrades as smooth as possible and with as much disruption as possible. Adherence to these standards ensures that the program is robust, invulnerable, and possesses the research design; indicating a flexible platform for future enhancements.

5.1.2 Hardware Standards

The hardware requirement for this research is aimed at helping to achieve a faster rate at which machine learning models, web page development, and data processes can easily be carried out. Resources include Google's Colab, which provides for an access to Tesla GPUs up to 16 GB of RAM as far as the training and evaluation of the models is concerned. For writing code and analyzing small amounts of data locally for development and testing purposes, personal computers with at least an Intel Core i5/AMD Ryzen 5, 8 GB RAM, and 512 GB SSD are recommended for fast code execution.A Web host, or the related hardware requirements entail the utilization of a dedicated or a virtual private Web host with not less than 4 CPU cores and 8 RAM bytes, in addition to proportional storage capacity as a result of increased numbers of users' interactions, data receipt, and analytical computation. Network infrastructure should ensure customers avail pertinent and faster internet connection for the exchange of real-time data besides being seamless on the website. These hardware standards ensure that a stable platform is created for the efficient implementation and deployment of the research project goals.

5.1.3 Communication Standards

The communication standards for this research project are developed to ensure that there is proper flow of communication through the activities of the team as well as exchange of data and coordination with other stakeholders. For normal business communication and updates, and regular emailing formalites such as – Gmail and similar instant messaging tools like slack are used. For business as well as project-related discussions and feedback sessions, it is necessary to schedule an on-screen meeting on Zoom or Google Meet. For the code share, most common and popular Version control tools like Git and GitHub is used for sharing code and tracking the changes to be handled in the project.Very common tools like Trello and Asana are used for the better understanding of the project and to keep the record of the tasks that are to be performed and the time-tables being followed. Any study data collected and documents prepared are stored and communicated, encrypted and through web tools such as google docs for ease as anonymity is maintained. These principles create productivity, organization and coordination among different teams which are helpful for the success of the project.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

It will combat the growing problem of drug addiction in Bangladesh with the goal to improve the lives of individuals and communities at scales never before achieved. By pointing out the main factors for addiction and creating a predictive model, the study allows early identification of high-risk subjects. By helping them find the cause of an addiction and affect treatment immediately, it also stops the insidious disease from likely progressing, reducing its harmful effect on theirs and their loved ones' health, emotional stability, and social connections. The study offers hope for families who learn early on that addiction poses risks and can provide proactive action for vulnerable people. This is able to help to lessen the emotional and financial load associated with this addiction, generally creating a more stable and a better family environment. By reducing the rates of drug addiction in a nice a round, the doing it leads to the creating of a safer built environment where such communities have less incidences of crime and violence connected with drugs. The results of the study can also inform programs aiming to create awareness and the choices to live healthy among youngsters. This research overall helps improve individual's and community life quality by means of the higher mental health, stronger social relations, more active life choices that in turn lead us to a more resilient and collaborative society.

5.2.2 Impact on Society & Environment

Specifically, the study covers growing drug addiction problem in a growing number of urbanized areas of Bangladesh. Machine learning is used to create a predictive model that allows early detection of at high-risk people and groups in addiction, thereby lowering addiction rates, and other social problems such as unemployment, crime and family problems. As a result, the study has changed public attitudes about addiction, presenting it as a preventable health condition, rather than a moral failing. It can simply help erase the stigma of addiction, encourage those stricken to get help, and offer a supportive recovery environment. Research provides insights, which can inform more targeted and more resource effective public health interventions to improve community wellbeing and foster harmony and resilience of communities, especially among vulnerable urban communities, through fighting the core causes and predicting factors of addiction. Secondly, it has a negative effect on the environment to the extent it 'helps' curtail actions that are threatening to the community like improper discarding of drug related trash which in turn pollutes soil and water sources. The study can help bring down the incidence of such risks, helping waste management systems, the environment, as addiction adds to them. By indirectly reducing the societal costs of addiction, the study mitigates the costs of unhealthy living and assists sustainable urbanization, freeing public resources from addiction and directing them toward environmental conservation and improvement activities.

5.2.3 Ethical Aspects

In studying it, the greatest ethical standards are maintained, especially when we are dealing with the chemically dense problem of drug addiction. Keeping participant privacy and confidentiality very crucial. We will take all reasonable precautions to ensure that all information provided is treated securely and the best technical and physical means will be taken to protect it against unauthorized access or disclosure. All participants will give informed consent on the goal of research, nature of data collected and their right to withdraw at any time. The objective of the study is to eradicate discrimination or bias in collecting and measuring the data to bring the results to the widest possible scope. Furthermore, the study's outcomes will be publicized in a manner that encourages neither stigma nor inappropriate assumptions about addiction or about vulnerable populations. In the form of a developed machine learning model for this study, preference would have been given to algorithms that provided interoperability because it is in order to avoid abuses such as profiling or unfair predicting of an individual based on a prediction. An alternative is to design the outcomes to spur public health actions and help build solutions. Research pursues this by following these ethical norms so as to ensure that the contribution to social health is positive and dignity and rights of all the participants are preserved.

5.2.4 Sustainability Plan

A detailed sustainability plan has been developed to ensure that this research has a long-term impact. The predictive model will next be updated with new data periodically, keeping it relevant and accurate in light of ever-changing socio-economic conditions and drug use patterns. The model will be improved over time through regular data collection in collaboration with local health care institutions, community organizations and public health departments. In terms of actual application of research results, the partnerships will be formed and partnerships will be made with government agencies, NGOs, and local community. The model will be implemented into public health programs and policy making through these collaborations as evidence-based addiction prevention and intervention initiatives. The results of the study are to also be communicated through workshops, seminars and training programs to stakeholders such as health care providers, social workers and community leaders on how predictive models can be applied to early detection and intervention. In addition, additional symbiotic funding sources, including grants from health-related organizations and collaborative networks with academic institutions, will also be used to support continued research and model maintenance. This long-term plan would keep project relevant and flexible, offer continued support in the battle against drug addiction and facilitate healthy communities.

5.3 Project Management and Financial Analysis

The project management and financial component of this research is very important to enable the successful implementation of the study within the specified budget and time frame. Effective project management involves numerous important aspects, including schedule planning, resource allocation, and risk management.

The project management framework is highlighted in figure 5.1, as is the list of activities necessary for the completion of the research and the time needed for each of them. The above figure maps each and every process of the project plan right from the planning stage of the project, design of the questionnaire to the acquisition phase, development of the model and reporting phase of the project. All assignments are mapped to the particular members of the team based on the specialization and responsibilities of every project. These are the managers of overall projects, researchers and data collectors and analysts together with the data scientists who are involved in the development of the models as well as the implementation of the models. Such separation of operations and duties guarantees that nobody relapses into amateurish execution of jobs, thereby improving effectiveness

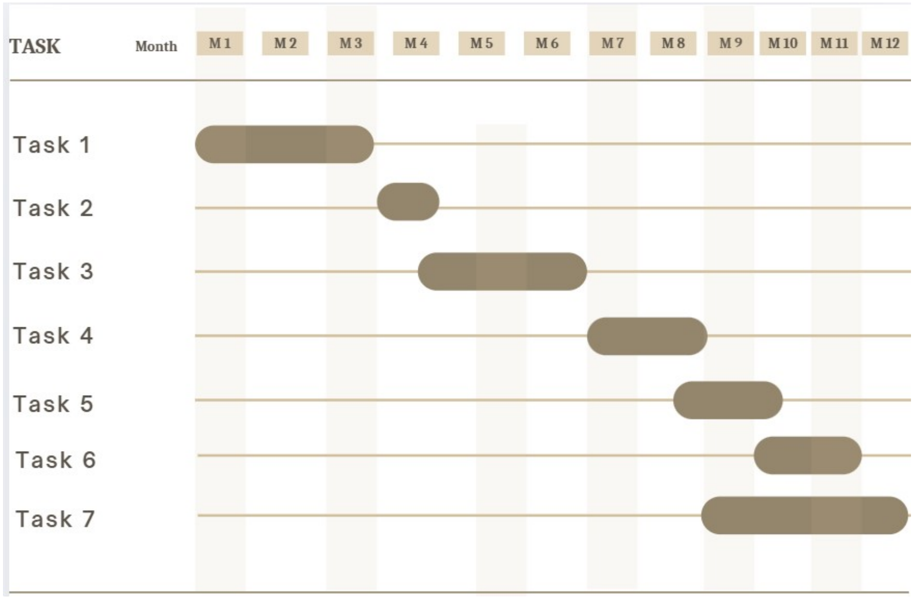


Figure 5.1: Project Management gantt chart

and quality. The total duration of the project estimated is 48 weeks and is broken down into clearly defined phases to make the work systematic and not simultaneous. Not only does such an approach guarantee fulfilling the set timelines but also the seamless transition between the distinct activities. Through this way, the project management plan ensures accountability by relating responsibilities to time horizons, embraces an efficient communication calendar and ensures that the resources are well managed. Finally, this much planned strategy enables the research to accomplish the set objectives in a way that is most effective in the given resources but at the same time, the study will afford quality and accuracy in the results as it seeks to meet its goals.

Table 5.1: Financial Analysis

SN	**Item**	**Estimated Cost (BDT)**
01	Questionnaire Design	5,000
02	Survey Distribution	10,000
03	Data Collection	20,000 – 30,000
04	Data Storage	5,000 - 8,000
05	Data Analysis Software	2,000 – 3,000
06	Hardware	80,000 – 90,000
07	Miscellaneous	10,000
Total= 1,32,000 – 1,61,000		

The estimated cost provides a breakdown of expenses related to various aspects of the research project. The costs include questionnaire design, survey distribution, data collection, and storage. While the data analysis software is open-source and free, computational resources for model training and testing, as well as ethical compliance costs, contribute to the overall budget. Miscellaneous expenses are included to cover unexpected costs.

Table 5.1 presents the costs estimates in relation to some of the parts of the project as described below. The questionnaire design is budgeted at 5,000 BDT to represent the cost in preparing a structured and efficient survey instrument. Survey distribution is provided a budget of 10000 BDT to cover the expenses incurred during the distribution of questionnaire among the participants. The data collection, including receipt of responses, is particularly expected to take between 20,000 and 30,000 BDT. Data storage have been estimated at 5000 to 8000 BDT to maintain proper security measures while dealing with the collected data. For the data analysis, it is expected to spend 2,000 to 3,000 BDT to purchase necessary software to make efficient processing and analysis of the dataset. Tangible needs, which include computing equipment, and is easily the highest cost, which lies between 80,000 to 90,000 BDT. Finally, miscellaneous costs are assumed to be 10,000 BDT as they embrace all incidental and other over head charges which cannot be easily quantified in advance. Total project cost is implemented in the range of 1,32,000 BDT to 1,61,000 BDT providing sufficient funding to cover necessary means for the project completion.

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

In this section, provide a mapping with problem solving categories. For each mapping add subsections to put rationale (Use Table 5.2). For P1, we need to put another mapping with Knowledge profile and rational thereof.

Table 5.2: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent of Stakeholder Involvement	EP7 Interdependence
√	√		√		√	√

Mapping with Knowledge Profile for EP1

This table 5.3 is designed to map the EP1 to the Knowledge Profile.

In order to attain Complex Engineering Problems (EP1) to facilitate the indicator of depth of knowledge, the specific Knowledge Profiles (K) need to be met. In the project,

Table 5.3: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
√	√	√	√	√

we have used Knowledge Profile K3: theoretical basis of engineering; K4: state of the art engineering specialist knowledge, for practical application; K5: engineering design; K6: engineering practice; and K8: research findings. As for the knowledge, we have previously learned in the course, we have been able to apply machine learning algorithms, software development life cycle, data compliance to actualize this project. In this way, we meet the demands of EP1. Furthermore, through defining and analyzing the issues regarding conflicting requirements, we have achieved EP2 to address the range of the conflict. For example, during the data collection process, inclusion of stakeholders was of essence. When we were struggling to determine the questions for data collection we have responded to the fourth element of the framework, understanding of the issues, and the sixth element – involvement of stakeholders and handling the conflicting needs. Finally, based on the concept of assembled components that make up this full project and the widespread use of different sorts of interconnected subsystems within this framework, it is possible to point to their complete interdependence in attaining the main goal of combating drug addiction. This integrated approach provides evidence to our capability to solve technologically challenging problems of engineering practice in a manner that optimally solves coding, practical, and stakeholder integrity issues.

5.4.2 Engineering Activities

In this section, provide a mapping with engineering activities. For each mapping add subsections to put rationale (Use Table 5.4).

Table 5.4: Mapping with complex engineering activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
√	√		√	√

However, apart from knowledge demander-oriented activities we have to perform several Engineering Activities (EA). For instance, to address EA1 (Range of Resources), we need a list of resources that include computational framework, computational resources, datasets, ethical principles, compliance procedures, and so on. These are the resources which are necessary for the right functioning and the successful undertaking of our project. In addition, EA2 (Level of Interaction) is realized during the collection of the dataset through physical surveys. Those are not just armchair activities but enable the accurate collection of data while supporting user engagement and trust. Furthermore, the project fits EA4, as its focus is on ethical information practices that enhance decision making, create real social values, and respect environment. By so doing, the project guarantees a comprehensive strategy in achieving its objectives and at the same time exercise high ethical and professional standard.

5.5 Summary

This project shows how to approach different kinds of Complex Engineering Problems (EP) and perform Engineering Activities (EA). EP1 is attained through the amalgamation of theoretical and the practical by use of engineering fundamentals (K3), specialized knowledge (K4), design principles (K5), practical applications (K6) as well as research findings (K8). Due to prior knowledge on the algorithms used in machine learning, software development processes and data compliance, these requirements have been met adequately. Struggles and issues faced when completing stakeholder engagement and dataset collection demonstrate EP2 (Range of Conflicting Requirements), EP4 (Issues Familiarity), and EP6 (Stakeholder Participation), which illustrate simultaneous capability to align to and manage epistemic complexity and collaboration. It also meets EA1 through bringing into play computational resources as the computational tools, machine learning frameworks, datasets, and ethical guidelines. The use of physical surveys to interact with the participants makes it meet EA2 and increases the user's trust with the system. CC a appropriate employ of informative synchronizations EA4 to facilitate proper decision-makers, societal gains, and environmental balance. collectively, these trends demonstrate an assimilated and moral systematic approach to attaining the overreaching objectives of the project.

Chapter 6

Conclusion

6.1 Summary

In this work we attempt to tackle the enhance drug dependance issue in Bangladesh by developing a predictive model using the machine learning techniques especially in our urban regions. The study learns through a mix of demographic, psychological and environmental data the influential aspects that make you addicted, even when you are not. This work establishes a prediction model that makes for a powerful tool for early detection of individuals and communities placed at high risk of addiction, allowing more targeted and successful therapy. The results highlight the influence on development of borderline addictive behavior by factors like peer pressure, economic level, family history and the drug accessibility. This work not only adds to addiction research but also provides valuable lessons for public health professionals and policy makers in Bangladesh using machine learning methods, such as logistic regression, decision trees and random forest. At the same time, this study emphasizes the need for model interpretability to facilitate the practical use of predictions in real world settings and to minimize stigma and encourage early action. Thus, the study lays the groundwork for future research in the prediction of the addition, specially in the similar socio-cultural context, and serves as a sustainable paradigm for the management of drugs addicted people using data based tactics. This study as a whole is a great step is addressing drug addiction in Bangladesh and moving towards a healthier and more supportive society.

6.2 Limitation

The scope of and generalizability of this study are limited by several limitations. Drug addiction in Bangladesh is largely under reported, owing to societal stigma and paucity of data infrastructure to gather data, and this is a key obstacle. Self reported data from surveys and questionnaires used in the survey do tend to have biases — under reporting and recalling of events inaccurately etc. Second, the cross-sectional design of study limits the establishment of causal links between such causes and addiction. Another drawback is that the sample population is not met as a focus on the urbanised areas may misrepresent patterns of addiction in more rural or remote locations. Most machine learning models are good at spotting trends and predicting risk, but success is dependent on the quality and variety of data. On the other hand, interpretability of complex models including ensemble approaches may become problematic to solve what finds to what actions. And finally, the study did not take into account severe genetic or neurological information, which are important traits in learning drug dependence but are exhaustive and specialized research. In the future study, these constraints could be addressed to dramatically improve the robustness and utility of the results.

6.3 Future Work

This study paves a way for exploring a range of opportunities in the area of addiction prediction and prevention especially in a third world country like Bangladesh and such. One development could be to further refine the model by using more various datasets, such as longitudinal studies which follow people over time to investigate the longitudinal nature of addiction as well as the long term impact of the treatment. The addition of genetic, neurological and even more complicated environmental factors increases the predictive model accuracy and can help better understand that addiction is a multidimensional phenomenon. Also, the study takes advantage of the architecture of the machine learning to test out sophisticated algorithms like deep learning and reinforcement learning being able to identify more complex patterns and improving prediction accuracy. One next study might explore for example the implementation of the notion on real time systems for rapid intervention in high-risk locations or individuals. A second area of research is for an application of the model to other addictive habits, such as alcohol and gambling, to determine whether the same risk factors hold. Therefore, the socio-cultural elements that influence the addiction to drugs in society should bear much deeper qualitative research in order to strengthen the scope of intervention and prevention. Such will help in designing region specific solutions, that are technically sound and socially acceptable and culturally sensitive.

References

[1] X. Zhao and e. al. Predictive analytics for youth substance abuse risk. *Addiction Science Clinical Practice*, 13(4):450–460, 2018.

[2] T. Nguyen and e. al. Deep learning for relapse prediction in addiction recovery. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:567–578, 2021.

[3] N. D. Volkow and e. al. Urban stress and genetic predisposition in addiction. *Nature Reviews Neuroscience*, 18(4):242–256, 2017.

[4] M. Hasan M. Rahman and S. Islam. Urbanization and substance abuse: A study on emerging trends in bangladesh. *Journal of Urban Health*, 97(3):345–356, 2020.

[5] W. H. Organization. Global status report on drug use and health. 2019.

[6] D. M. Paulus and J. M. Tapert. Adolescent brain development and the risk for alcohol and other drug problems. *Neuropsychology Review*, 28(2):203–220, 2018.

[7] A. Khan and e. al. Socioeconomic determinants of drug addiction in south asia. *Addiction Research Theory*, 26(2):112–120, 2018.

[8] W. K. Bickel and e. al. Reinforcement learning models in addiction. *Behavioral Neuroscience*, 133(3):200–212, 2019.

[9] U. E. Ghitza and e. al. Personalized treatment approaches using predictive analytics. *Substance Abuse*, 34(2):105–114, 2021.

[10] H. K. Mehta and S. Chatterjee. Machine learning applications for behavioral health outcomes in addiction recovery. *Journal of Biomedical Informatics*, 112:103607, 2020.

[11] A. Kumar and e. al. Socio-demographic data integration in machine learning frameworks. *Addictive Behaviors*, 95:153–164, 2019.

[12] Y. Xie and e. al. Interpretable ml for opioid addiction prediction. *Interpretable ML for opioid addiction prediction*, page 104, 2020.

[13] Z. C. Lipton. The mythos of model interpretability. *Communications of the ACM*, 61(10):36–43, 2018.

[14] A. J. King and E. T. Fennel. Economic impacts on addiction recovery: Insights from low-income communities. *International Journal of Social Psychiatry*, 67(3):215–224, 2021.

[15] A. Nair and e. al. Regional variations in addiction prevalence. *Social Psychiatry and Psychiatric Epidemiology*, 55(8):1059–1070, 2020.

[16] P. Gupta and M. Patel. Exploring urban stress as a predictor of substance abuse trends in south asia. *Asian Journal of Psychiatry*, 49:101993, 2020.

[17] S. Ahmed and e. al. Family history and addiction risks. *Journal of Studies on Alcohol and Drug*, 79(6):868–875, 2018.

[18] Y. Zhang and e. al. Enhancing addiction models with qualitative data. *Health Data Science*, 2, 2019.

[19] S. Lee and e. al. Integrating qualitative and quantitative predictors in addiction research. *Journal of Behavioral Data Science*, 4(2):35–48, 2020.

[20] R. Sultana and e. al. Drug accessibility in urban bangladesh. *Bangladesh Journal of Public Health*, 31(1):22–34, 2021.

Predicting Drug Addiction Patterns in Urbanizing Bangladesh: A Machine Learning Approach

ORIGINALITY REPORT

SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
14%	10%	6%	9%

PRIMARY SOURCES

1	Submitted to United International University Student Paper	4%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	Submitted to Daffodil International University Student Paper	2%
4	Submitted to Grambling State University Student Paper	1%
5	de Carvalho Pereira Ferreira, Inês. "Churn Prediction in Digital Marketing", Universidade do Porto (Portugal), 2024 Publication	<1%
6	www.mdpi.com Internet Source	<1%
7	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	<1%