A CLASSIFICATION BASED MACHINE LEARNING MODEL TO PREDICT SUICIDAL THOUGHTS: BANGLADESH PERSPECTIVE

BY

MUNIRA FERDOUS
ID: 172-15-9772

AND

JUI DEBNATH
ID: 172-15-9712

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

Supervised By

Narayan Ranjan Chakraborty
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JUNE 2021

APPROVAL

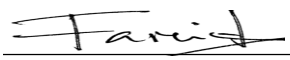
This research project titled "**A classification based machine learning model to predict suicidal thoughts: Bangladesh Perspective**", submitted by Munira Ferdous, ID: 172-15-9772 and Jui Debnath, ID: 172-15-9712 to the Department of Computer Science and Engineering, Daffodil International University 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 3 June 2021.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Gazi Zahirul Islam — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Raja Tariqul Hasan Tusher — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Dewan Md. Farid — **External Examiner**
Associate Professor
Department of Computer Science and Engineering
United International University

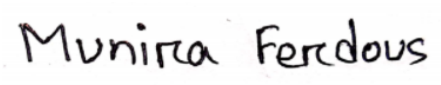
DECLARATION

We hereby declare that this project has been done by us under the supervision of **Narayan Ranjan Chakraborty, Assistant Professor**, Department of CSE, 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.

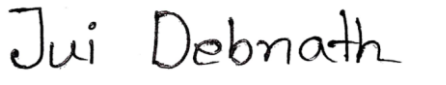
Supervised by:

Narayan Ranjan Chakraborty
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Submitted by:



Munira Ferdous
ID: 172-15-9772
Department of CSE
Daffodil International University



Jui Debnath
ID: 172-15-9712
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

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

We would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan**, Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of the CSE department of Daffodil International University.

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

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

ABSTRACT

Suicide is an unnatural death and it becomes a major issue in Bangladesh. The World health organization report which is published in 2018, it says, Suicide Deaths in Bangladesh come across 9,544 deaths, or in percentage, it is 1.23% of entire deaths [1]. We aimed to develop a model that can predict suicidal thoughts, using a machine learning algorithm. It can prevent future risk of suicidal attempts. Dataset presents 15-46 years old people's thoughts, feelings, their regular activities, contains a total of 22 attributes and 441 instances. The classification process was performed using nine machine learning algorithms those are, Naive Bayes, KNN, Linear SVC (support vector classifier), Non-linear SVC, Random Forest Classifier (RFC), Decision Tree, Logistic Regression (LR), and Extreme Gradient Boosting (XGB) Classifier, Adaptive Boosting(Ada-boost) Classifier. The prediction model achieved a good performance. The highest accuracy achieved Random Forest Classifier (0.91). The area under the receiver operating characteristic curve (AUC)=0.9 for Random Forest Classifier. This study shows the probability that a machine learning approach can able to decrease suicide risk. Hopefully, this model will assist as a support for reducing future suicidal risk. The paper ends with a review of various practical issues, which may be explored to enhance model performance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

KNN = K-Nearest Neighbors

CSSRS = Columbia Suicide Severity Rating Scale

BSS = Beck Suicide Ideation Scale

BPNN = Back Propagation Neural Network

EHR = Electronic Health Record

GSHS = Global School-based Student Health Survey

XG Boost = Extreme gradient boosting

Ada Boost = Adaptive Boosting

CHAPTER 1

INTRODUCTION

1.1 Introduction

Suicide is now-a-days a very common social taboo. It is known as unnatural death. In Bangladesh, it is a common reason for unnatural mortality. Due to suicide, worldwide many people died every year. An article shows, among them 2.06% of people are Bangladeshi [2]. There are many reasons behind suicidal deaths. Social media bullying, depression, loneliness, Academic frustration, humiliation, and many more reason behind suicide. The COVID-19 virus has infected and died many people in today's pandemic situation. Coronavirus Death is converging international barriers and normally attacking a wide number of people. But a foundation which is named Alcohol Foundation shows that several people in Bangladesh died from suicide in the last year than those who died on the Covid-19 virus. A study shows Bangladesh recorded 70% more life losses from self-destruction than COVID-19 [3].

Among adolescents who have an age range between 15 to 24 years, 5.5 percent of girl's children and 4.8 percent of boys did a suicide attempt [4]. Most of those who attempt suicide are adolescents in Bangladesh. But a matter of fact that the teenage suicidal tendency and behavior is ignored most of the time in Bangladesh and many other low-income countries. Amongst teenagers, the predominance of lifetime suicidal thoughts is moderately high. Based on a pure dataset, it is possible to predict whether some people are thinking suicidal thoughts or not.

There is some research work that has been done about suicidal thoughts prediction, also different types of suicidal behavioral problem prediction such as depression, drug, and much more related research work have been done previously. At this time, we have tried to build a prediction model built on machine learning. Based on the dataset, among those who have suicidal thoughts, this paper also showed the percentage of people who have taken a drug, suffer from loneliness, and whether they have any depression.

1.2 Motivation

Machine learning models can detect or predict future data from the past data. This permits them to continuously improve predictions based on the new output and various data and provide an automated result. This is how the machine learning-based model works. Machine learning algorithms can predict the algorithm's accuracy.

Suicide is a big social issue. Day by day it is increasing. In the pandemic many people died from Coronavirus. But a report shows more people die from suicidal ideation than from the coronavirus. Most of the reasons for suicidal attempts are a mental disorder, social media trolling, bullying, financial problems etc.

A report shows, During COVID-19 restraints, many youngsters and teenagers used the internet and social media to attach with one other. Some became entangled in serious implicit relationships as they spent a long time online. Many are mentally depressed due to many social taboos and illicit relationships and from there take suicidal attempts.

There is growing evidence of using Facebook some young adults express their suicidal thoughts and emotions, predominantly shared in social media. We examined most of the sufferers were teenagers and young adults. They are reluctant to share their self-destructive or harmful thoughts with their physicians or other family members.

To make awareness of sufferers parents and well-wishers we build a model where they easily predict if their child has any suicidal thoughts. If they know about it, they would have taken special care of these children. Similarly, psychologists are also helpful from the model. They also learn about if their patients have any suicidal thoughts or if they are in mentally depression or taking any drugs or not.

The young generation is our country's future. If they die prematurely, it will be unfortunate for our country. If it is possible to get them back on track with some special care, we are developing a model through machine learning to find out if anyone has any suicidal tendencies. Parents, psychologists and everyone will learn the model about their family member’s/friends behavior and the model will predict if their children have any suicidal thoughts or not.

1.3 Rationale of the Study

Drugs, depression, suicide destroyed human life. It is not limited to the person who commits suicide but it has become detrimental to society. According to the police report, approximately 7,671 women died by suicide in the year of 2012 to 2017. Between them, 3,444 occurrences happen at their parents' houses and 3,927 incidents happen at their in-laws' house. In opposition, about 9,212 men were sufferers of suicidal death. A study shows, aged women had depression and loneliness was the principal factor [5].

Among more adolescent children, suicide thoughts are often unpredictable. They may be connected with emotions of sadness, nervousness, anger, or difficulties with concentration and hyperactivity.

Among teenagers, suicide attempts are connected with their emotions of anxiety, self-doubt, stress to succeed, financial insufficiency, frustration, and loss. Teens thoughts there is a solution, that is suicide.

Especially for a teenager who is in a state of nervousness, when his parents or his family members are in the eye of the beholder, they can't do anything. some of the teenagers don't want to share all things with their parents. the physicist also could not find out the actual problem of those teenagers.

Machine learning can build and predict a model. We took some people's data from a dataset. After building a machine learning-based model It gives a prediction of teens and adults and every person whose data is given as input if they have any suicidal thoughts. How many people take drugs and are in mental depression also be shown in this paper based on our dataset. This model helps to predict. someone's suicidal thoughts. Psychiatrists and psychologists are also able to understand if their patients have any mental disorders Suicidal thoughts and Depressions are preventable mental disorders. If someone especially parents and doctors is known about their behavior they will be able to bring them back to the right path.

1.4 Research Questions

- What does it mean by suicidal thoughts?

- Causes of suicidal thoughts?
- Difference between Suicidal thoughts and suicidal attempts?
- Symptoms of suicidal behavior?
- How does the machine learning model work?
- What is the best accuracy algorithm?
- How to pre-process the dataset?
- What are the future works of this model?
- How does the suicidal thoughts prediction model work?
- What is the advantage of this model?
- Should we utilize the most commonly used machine learning algorithms?

1.5 Expected Outcome

This is a research project, our main interest was to publish a research paper in a relevant area. Research works as a perpetual process. Many researchers do investigate particular research questions for searching for an effective solution. Maximum researcher research about depression, drug prediction of university students, different mental disorder predictions based on machine learning has been done. Some research has been done on what percentage of suicide happens every year or how it can be predicted. Our research motive is to aware parents and helps doctors to provide the best accuracy-based model. So that our young teenagers will aware of committing suicide, taking drugs, and depression. There is a way to do a machine learning-based model with the best accuracy. In this research, we introduced a machine learning-based model to keep general people from committing suicide.

1.6 Report Layout

This entire report contains total of 6 chapter. Total 6 chapter has been coved with some subsections. At first, chapter 1 covers an overview of the whole research work such as, the research motivation, introduction, research question, expected outcome etc. Chapter 2 covers the background study of the research. In chapter 3 this paper contains the methodology of the

research. Chapter 4 covers the experimental result analysis. In Chapter 5 shows, the impacts of society, environment and sustainability with some subsections. The last chapter, chapter 6 covers research conclusion and future works.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

The machine learning approaches is a way to develop an automatic process. Defining the problem appropriately in the first step in building a machine learning model is data collection and goal defining. The machine learning process and ready to model. Estimates suggest that the death toll could rise to 1.5 million by 2020. Recently there has been much concern to explore suicide among young people [6]. A two-stage study was conducted to honor suicidal ideation. It was found that the average life expectancy of suicidal ideation was 5 percent in young adults. The majority of young people with suicidal thoughts were women 52.8%, single 82.4%, and students 92.73% [7]. They take drugs and many dangerous substances because of their harmful effects. Many intelligent students lose their lives each year as a result of suicide. It is manageable. The rate goes up a lot and you have to stop with the right step. Psychiatrists often help teens deal with depression and all suicidal ideation and suicidal thoughts. Too many people who submit submissions do not talk about their weaknesses. But their behavior is altered by their suicidal thoughts and most of them commit suicide. Many parents are unaware of their children's condition. We are focused on the changing behavior of young people and adults. If parents become aware of this behavior, they will become aware of their children's attitudes with the help of our machine-based suicide prediction model. Drugs have affected the younger generation. Drugs, depression are reasons why young people and adults have suicidal reasons. It can also be a method of learning predicted machines. This machine-based model helps teens and teens avoid drug addiction, depression, and suicide. This study helps to make parents aware of their children. Also, psychiatrists and psychologists will help in this way to better manage their patients.

This research would help to get rid of their depression and help to lead a normal healthy life.

2.2 Related Works

Suicide is a serious issue, notably among young adolescents and mixed ages people that can have enduring harmful impacts on individuals, families, and societies. Many researchers were

researched suicide. There is much research about this issue based on machine learning prediction models.

Some researchers intend to address the scarcity of terminological sources associated with suicide by a method of assembling connected with suicide. For a better analysis, this study also proposes, to examine Weka as a tool of data mining. They completed the research based on machine learning algorithms that can obtain beneficial data from Twitter data accumulated by Twitter. The result of Cross-validation of evaluations on classifiers for assumed tweets with risk of suicide naive bias 87.50% [8]. A study classified suicide attempters for those troubled by schizophrenia. They apply sociocultural and clinical features based on the machine learning approach. They conveyed a cross-sectional estimation on a representation of 345 members diagnosed who are with schizophrenia spectrum disorders. They recognized Suicide attempters and non-attempters using the CSSRS also they use the BSS. They used four classification algorithms to train the models. After training their dataset they got the highest accuracy 67% from the Support vector machine [9]. Some of the researches research about Prediction models for the huge risk of suicide in Korean adolescents utilizing machine learning methods. This study aimed to improve the prediction model to identify Korean adolescents of high-risk suicide utilizing machine learning techniques. They used a nationally representational dataset of the Korea Youth Risk Behavior Web-based Survey. They utilized five machine learning algorithms and achieved the best accuracy for XGB which is 79.0% [10]. One study paper aimed to generate a model predicting individuals with suicide ideation utilizing a machine learning algorithm. The prediction model achieved a better performance (AUC)=0.85 [11]. A Study predicts future risk of suicidal ideation based on a machine learning approach. They collect data from Twitter data. They trained a series of neural networks on their collected dataset toward suicide-related psychological constructs. Trained a random forest model applying neural network result for predicting binary Suicidal ideation state. This research achieved an AUC of 0.88% [12]. A research paper proposed to examine the potential of machine learning to predict future suicidal behavior utilizing population-based longitudinal data. This paper achieved the random forest algorithm was the best algorithm to predict suicide ideation. And gradient boosting is the best to predict suicide attempts [13]. A study attempted to evaluate Critical risk factors for suicide attempts and death. The result they got for every one of the five symptoms was raised in those involved in suicidal behavior ($p < 0.05$) [14]. One survey paper did A

Systematic Review and Simulation to estimate the diagnostic efficiency of suicide prediction models in predicting suicide and suicide attempts. Reviewed 7306 abstracts,17 group studies, serving 64 unique prediction models then the survey achieved the result is Global classification accuracy is better [15]. A study build a prediction model for a suicide attempt based on BPNN. After implementation, the research achieved the specificity (93.9%), sensitivity is (67.6%), negative predictive value (84.1%), positive predictive value (86.0%) [16]. A research paper's idea was to Predicting Suicidal Behavior from longitudinal historical data, usually available in EHR methods. That can be used to predict patients' future risk of suicidal behavior. After training the dataset, the model performed 33%–45% of sensitivity and 90%-95% specificity [17]. In Bangladesh, there is very little work done. A Systematic Review estimate the diagnostic accuracy of suicide prediction models. This paper surveys predicting suicide and suicide attempts also simulating the results of performing suicide prediction models utilizing population-level evaluations of suicide rates [18].

The present time has been overlooked by the COVID-19 pandemic. Lots of people died from this virus. People are regretting some non-death-related situations. Maybe they lost a job, feelings of depression, and many other problems in their life. These causes are responsible for suicidal ideation. An article intends to review the present research on COVID-19 related suicides. The research paper involved anxiety in suicide survivors and highlight modifiable risk factors for both conditions, also this study paper shows, the impact of the pandemic circumstances on self-inflicted injury and its importance on families, friends and the Community [19]. Depression, drug addiction are the leading causes of attempting suicide among university students to adolescents. A study examined the role of subject-selection purposes and learning context circumstances in students' depression and suicidality. This study examined 960 undergraduates from five different Bangladeshi universities utilizing inquiries. The paper concerning almost half of the participants was depressed the percentage is 47.7% [20]. Again, another study design of the existing study was to examine the relationship between their emotional problems among university students. The research study representation is included in 112 University students, where females (61) and males (51) are represented, using the purposive sampling method [21]. Cross-sectional research was taken out between 1844 students registered at the University of Dhaka, Bangladesh. They performed Hierarchical regression analyses to examine the analytical power of the variables predicting depression in

this population. This study achieved the result of Depression predominance was 28.7% [22]. This study used data from the 2014 GSHS, Bangladesh. They examined the Risk factors of suicidal behavior. Used a generalized estimating equation-modified Poisson regression approach [23]. This study intended to examine the prevalence of and the factors connected with suicidal behavior between school-going adolescents in Bangladesh. The result was, age-adjusted ubiquity of suicidal behavior among adolescents in Bangladesh was 11.7% [24]. A research objective is to examine whether depression can be strongly predicted. The researchers aim to recognize depression in its early steps and secure a fast cure for sufferers so that unbearable occurrences such as suicide can be avoided. Used three algorithms to predict depression in university students. They achieved the highest accuracy for Random forest 75% [25]. This study was carried to add to the currently inadequate data by describing the pervasiveness of depression, describing sleeping patterns & suicidal tendencies amongst medical students. Researchers have reviewed the overall mental health situation among medical students in Bangladesh. They have shown the Suicidal tendency of medical students with a relationship with family and friends and their sleeping patterns [26].

In Bangladesh suicide is increasing day by day. Suicide is preventable if someone knows in advance that someone has suicidal thoughts. We aimed to develop a model predicting suicidal thoughts using a machine learning algorithm. The objective of our research is to prevent suicide.

2.3 Research Summary

Machine learning is a very popular approach for prediction and detection. There is much work already performed about prediction and detection with the machine learning algorithm and data mining process. Different researchers have different model build techniques.

Suicide is increasing day by day. It is preventable if everybody's aware of the topic. To know if someone has suicidal thoughts, we build a model based on a classifier algorithm that can predict suicidal thoughts. We built the model based on a dataset collected from 'Kaggle'. Furthermore, the dataset has been modified by means of SMOTE technique to address the imbalance difficulties.

Afterward, the remaining pre-processing tasks, such as features selection and data normalization. We trained the model out training data after a train-test split was made on the ratio of 80:20. We applied nine classifiers. We used logistic regression, decision tree classifier, random forest classifier. After declaring and defining all functions and libraries we trained the model and got a better accuracy for the model which is Random forest and the accuracy was 91%. RF gives higher accuracy during cross-validation.

In the random forest, It turns multiple trees into a model. Each tree gives a classification to classify a new object based on different attributes. In general words, Random Forest creates multiple trees and connects them to get a more accurate result.

While creating random trees it is divided into different nodes. Then it examines the best result from the arbitrary subsets. This results in a better model of the algorithm.

2.4 Scope of the Problem

Every year millions of people die as a result of suicide. In Bangladesh, some of the main reasons for suicide is depression, loneliness, other family problem, etc. It is increasing day by day. The thoughts of suicide come to mind before people attempt suicide. If we know in advance whether there is anyone have any suicidal tendency/thoughts in the cause, then suicide can be prevented a lot.

Our research aim is mainly to build a model by analyzing data for prediction and applying machine-learning algorithms. Our proposed model can predict suicidal thoughts. The young generation is staying with drugs, alcohol, smoking, and any other harmful things. Many people couldn't express their feelings properly, as a result, they get depressed and they thoughts about suicide. It destroyed our society. We build a prediction model that can easily predict a person's suicidal thoughts. Parents are coming to know about their children's activities. Psychologists, the psychiatrist will be making better treatment to their patients which could stop the suicide and make them motivated. This model can save someone's life. Machine learning and artificial intelligence tools are used for several object detection and predictions. Therefore, we chose that using machine learning, we would create a model prediction of suicidal thoughts. We are hopeful, It brings a positive impact on society.

2.5 Challenges

Very little work has been done in Bangladesh to predict suicidal thoughts. For us, it took a long time to understand the predictions of suicidal thoughts. People feel free to talk about this issue so it is very difficult to collect data. We searched the data for our research but faced some challenges. Coronavirus or COVID-19 is a highly contagious disease transmitted by newly discovered coronavirus. Most people are infected with the COVID-19 virus. Government announces lock worldwide. Bangladesh's situation is deteriorating rapidly. The government has announced the closure of the entire country. Due to this epidemic, we were unable to download our desired data. So we searched on online platforms such as, UCI, Kaggle for information about suicidal thoughts and what we wanted. We found a database almost identical to our findings. From there we selected some data as we liked and used it. We were unaware of any new machine learning tools and algorithms. By doing some tutorials about machine learning and the help of our manager then we take the whole implementation process. Also, we had to use a lot of algorithms to get the best model accuracy. This proved to be a real challenge for us.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This research paper followed a proper methodology to complete our research. In chapter 3, we are going to discuss the entire methodology of the research work. This section included a detailed discussion of using the machine learning approach with a short explanation of every part of the methodology.

The objectives of this thesis are to build a suicidal thoughts prediction model and a comparative analysis among classifier algorithms to choose the best out of them. In our research, we have used supervised learning classifier techniques to get the best accuracy for this predicting model. First, we have collected a dataset that contains all of our necessary data for our research. Every machine learning model needs a good dataset to find an accurate automatic system. Every machine learning model required a better dataset to build an accurate automated system. To create this model, we have used various classification algorithms. We applied Logistic Regression, Decision Tree, Random Forest, Naive Bayes (Gaussian), KNN, AdaBoost Classifier, XGBoost Classifier, Support Vector Machine (Linear), Support Vector Machine (Kernel trick/Non-linear/rbf) classifiers in this analysis.

22 primary factors were shown to be strongly linked to suicidal thoughts. We explored some of the circumstances according to the finding. We prepared our dataset before implementation. To choose the right algorithm for the model, we measured and calculated accuracy, sensitivity, specificity, recall, F1-measure/score, and roc-curve for each algorithm. We find the most reliable and sufficient random forest (91%) for our model. Numerical equations and graphical figures of the model have been illustrated in this report. This record is encouraging to learn the whole work.

We consider the complete research work as a framework. In the chapter, every actions of the research methodology are shortly presented. There are many subsections and core sections that help to understand the significance of the model. In this paper, we have also shown the

workflow of this research which gives a short view of how we completed the entire research work. After implementation, we found a random forest algorithm had the best accuracy.

3.2 Research Subject and Instrumentation

We are building a classification algorithm-based prediction model with the best accuracy found in a comparative analysis among 9 supervised classifiers. Our thesis topic name is 'A classification-based machine learning approach to predict suicidal thoughts in the context of Bangladesh'.

We have discussed the process of building a prediction model from the very first to the end of the research. Now-a-days, machine learning approaches are the most popular to predict data. In this research, we used 'Python' with its ML libraries, for instance, Scikit-Learn, Pandas, Numpy. The model is built in 'Google Colab' which is very convenient for its friendly UI/UX design. 'Microsoft Excel' and 'Pandas' were used to store and pre-process our dataset for this entire research work.

A listing is presented subsequently of the required tools for this pipeline of workflow.

Hardware and Software:

- Intel Core i5 8th generation
- Primary Storage: 8GB
- Secondary Storage: 1 TB
- Integrated Development Environment: Google Colab

Development Tools:

- Operating System: Windows 10 64-bit
- Python 3.7
- Microsoft Excel

Required library:

- Numpy
- Pandas

- Scikit-Learn
- Plotly
- Seaborn

3.3 Data Collection Process

According to our research purpose we have collected a raw and useful dataset from Kaggle. The raw data were based on people of different ages in Bangladesh using Google Form and handouts. After that, Statistical techniques like SMOTE have been applied since the original dataset had minority instance issues that could lead us to biased classification results. The raw dataset contains 26 attributes and 212 instances. After SMOTED the whole dataset, the dataset contains 22 attributes and 441 instances. Data pre-processing has been done using SMOTE technique so that after applying machine learning algorithm we will find out the best accuracy algorithm for the model. SMOTE is known as, synthetic minority oversampling technique.

It is one of the most usually used oversampling techniques to determine the imbalance difficulties. It works for balancing the class distribution by randomly building minority classes. The SMOTE class seems like a data reconstruction object from Scikit-learn. It must be determined and configured, it fits at a dataset, then utilized to build a new modified version of the dataset.

To reached our goal, we have searched for collected our data based on the following feature:

- Age
- Gender
- Education level
- Thoughts or feelings about drug
- Their most spending time with
- Failure in their life
- Emotional problem
- Suicidal thoughts
- Family relationship
- Financials of family

- Addicted person in family
- Withdrawal symptoms
- Satisfied with workplace
- Is they have any case in court
- Living with drug user
- Smoking
- Ever taken drug
- Friends influence
- If chance given to taste drugs
- Easy to control use of drug

3.3.1 Data pre-processing

We prepared our dataset before implementation. To normalize the certain columns of data used the standardize by standard deviation technique function. This dataset has normalized for certain columns of the values in the dataset.

The formula for normalized value is = $(x – x) / s$

Here,

- x = Value of the data.
- x = dataset mean value.
- s = standard deviation of dataset

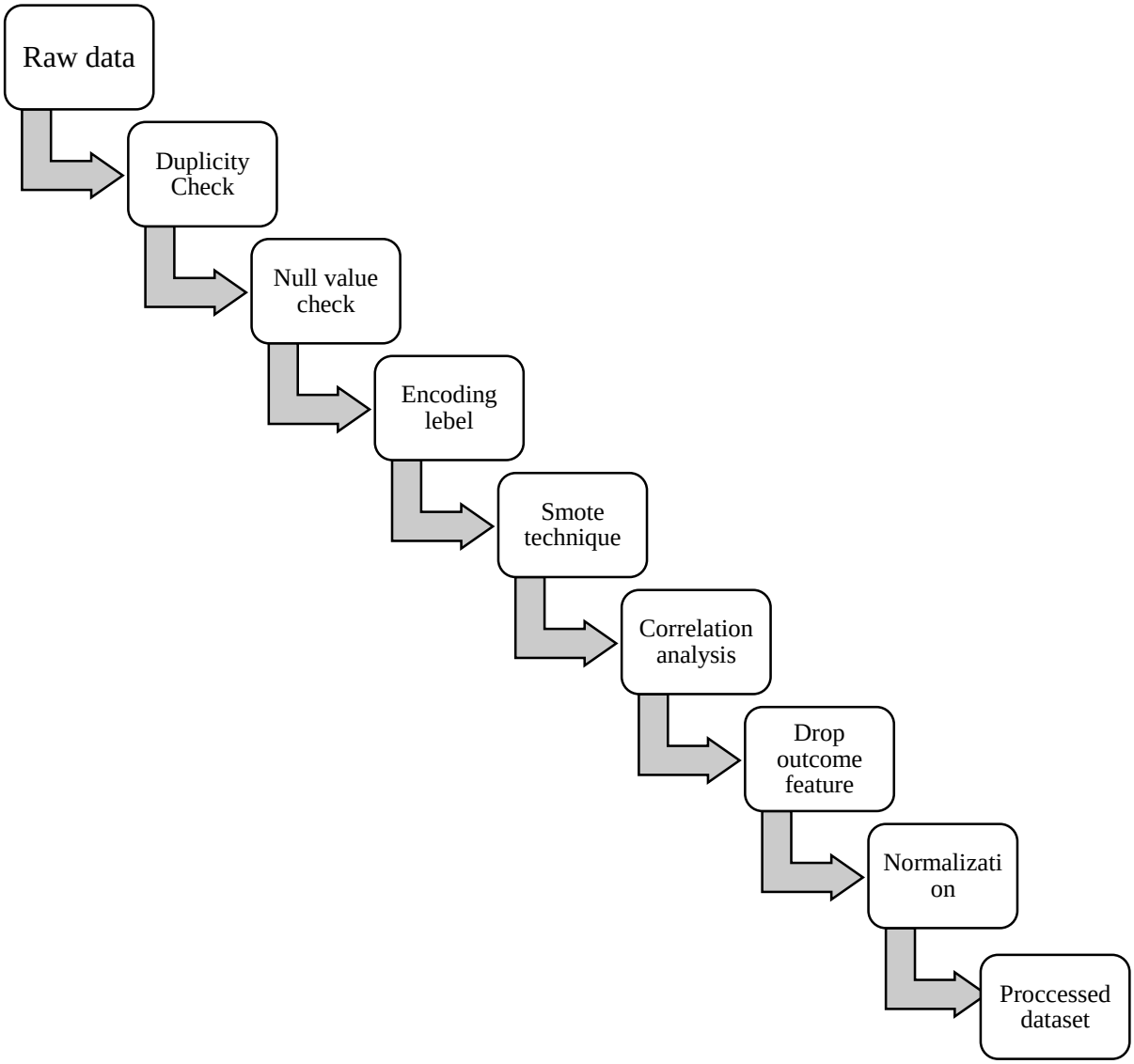


Figure 3.1: Steps of Data Pre-processing

3.4 Statistical Analysis

3.4.1 Correlation of 10 attributes with Suicidal thoughts

This section shows the co-relation of 10 attributes with their target variable suicidal thoughts. How many people have suicidal thoughts and how many have not, that is calculated in table 3.1

TABLE 3.1: CORRELATION OF 10 ATTRIBUTES WITH SUICIDAL THOUGHTS

Variables		Number of Yes Rate	Number of No Rate
Live with	Hostel/Hall	42	68
	With Family/Relatives	73	141
Motive about drug	Should avoid	4	7
	Social trend	81	113
	Disease	22	83
Spend most time	Alone	8	45
	Family/ Relatives	84	152
	Friends	15	19
Failure in life	Yes	93	25
	No	105	113
Mental/emotional problem	None	54	69
	Anger	47	77
	Depression	14	28
	Tension	0	3
	Others	1	38
Financials of family	Poor / weak	29	8
	Medium	53	104
	Rich / Strong	18	45
	Solvent	10	40
Living with drug user	Yes	49	96
	No	53	109
	Not sure	11	1
Smoking	Yes, every day	57	127
	Yes, occasionally	40	36
	No, I don't	17	55
Ever taken drug	Yes	79	127
	No	38	87
Friends influence	Yes	64	89
	No	54	119

3.4.2 Percentage of Correlation

This portion is calculated by the formula,

Correlation of Depression with Suicidal thoughts

Let, n = Number of Depression

N = Number of Suicidal Thought

n = 14, N = 126,

So, n/N*100 = 11.1%

And

n = Number of No Depression

N = Number of No Suicidal Thought

n = 28, N = 236,

So, n/N*100 = 11.9%

By the similar formula figure 3.2, 3.3, 3.4, 3.5, 3.6, 3.7 is created.

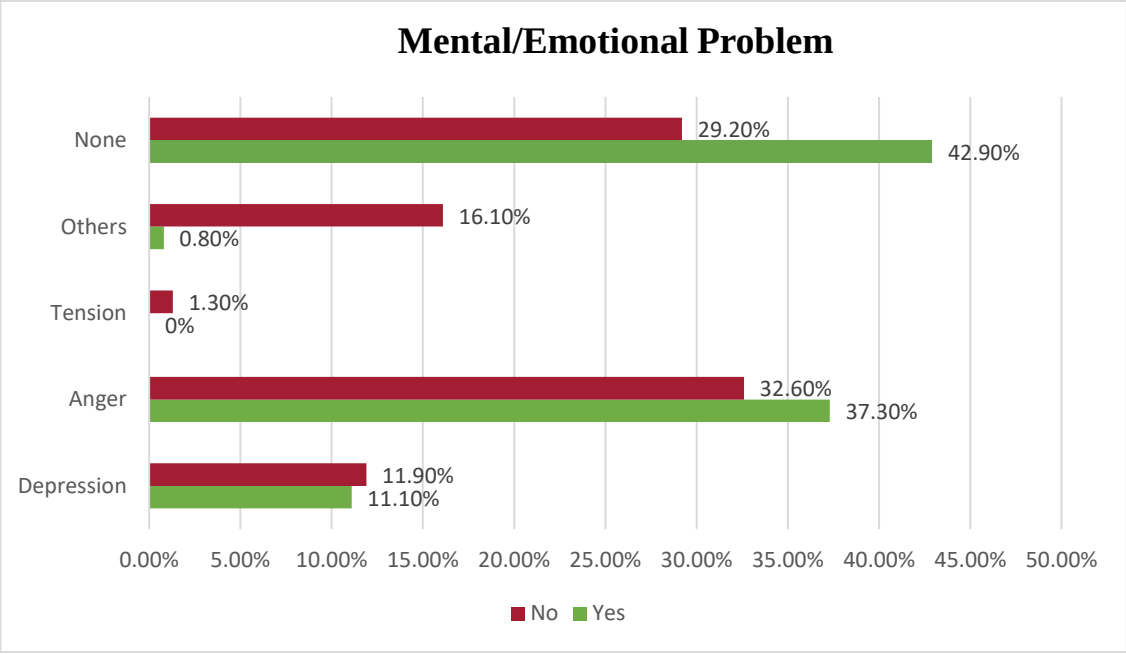


Figure 3.2: Mental/Emotional Problem for Suicidal Thoughts

Figure 3.2 shows that, percentages of people who have suicidal thoughts or not due to depression, anger, tension and other problems.

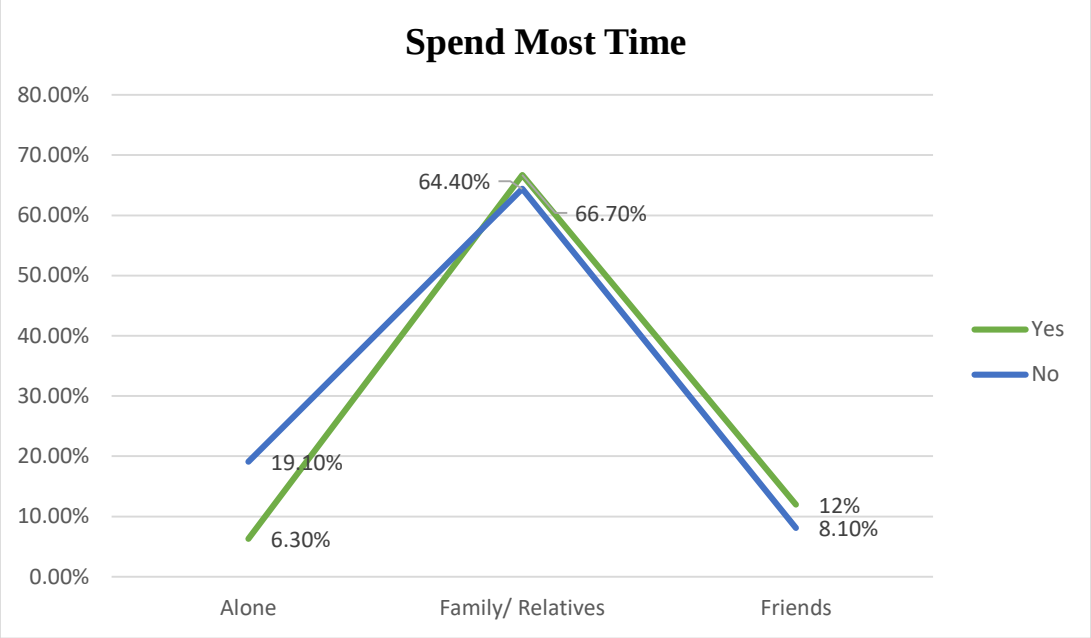


Figure 3.3: Effect of Time Spending

Figure 3.3 displays, having suicidal thoughts persons how they spend their time. We calculate the figure from our dataset.

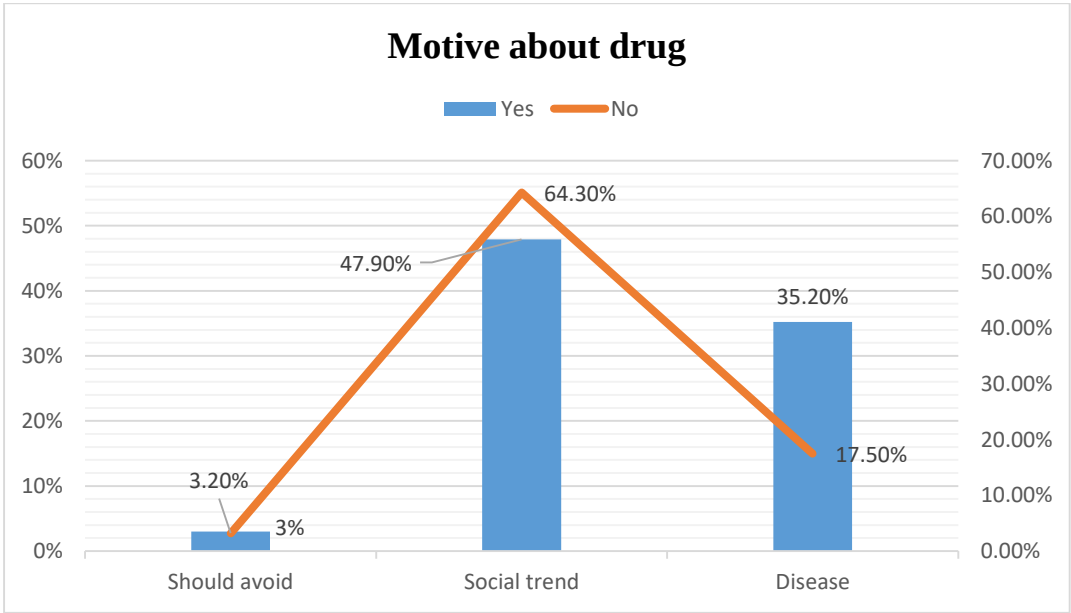


Figure 3.4: Impact of Motive about drug

Figure 3.4 shows, those people who have suicidal thoughts already, how many of them have motive about taking drug.

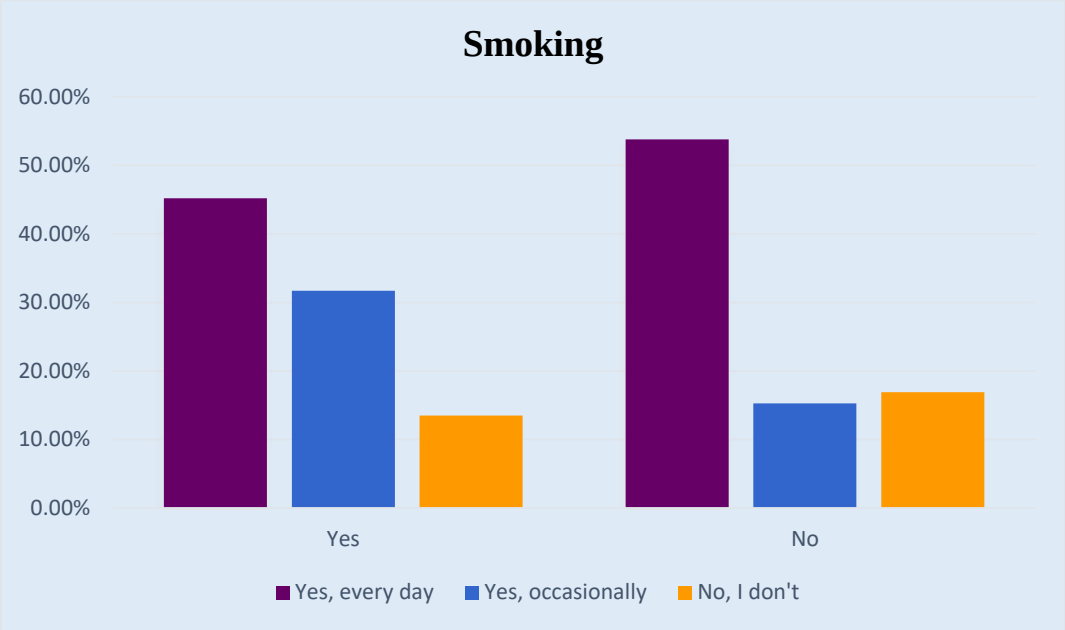


Figure 3.5: Suicidal Thoughts Rate Based on Smoking Nature

Figure 3.5 shows the percentage of doing smoke.

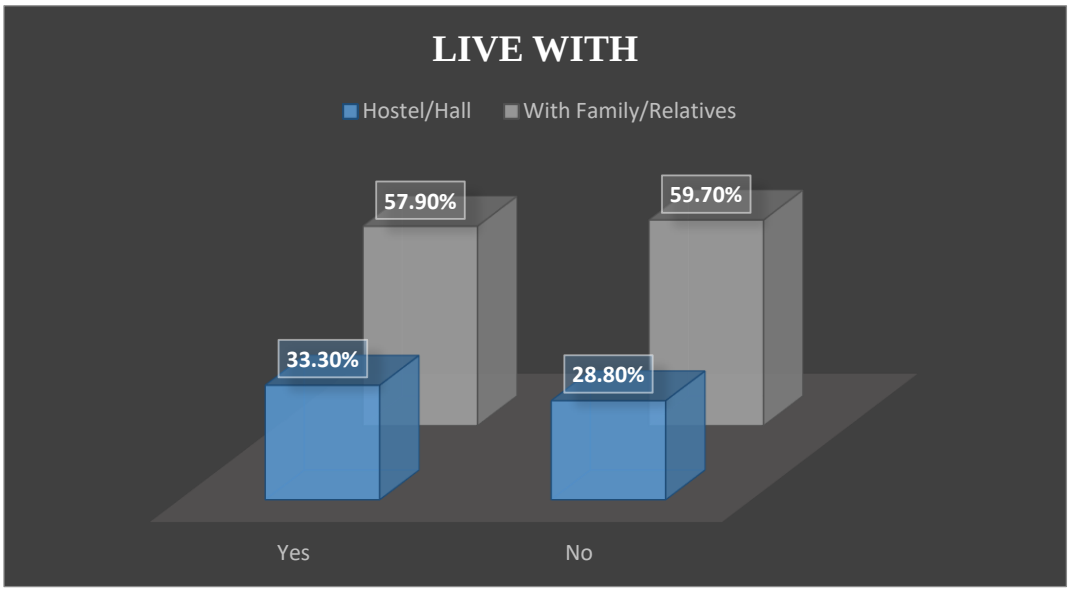


Figure 3.6: Living Nature

Figure 3.6 shows, where they love most of the time. Figure 3.7 shows the financial condition of those person who have suicidal thoughts.

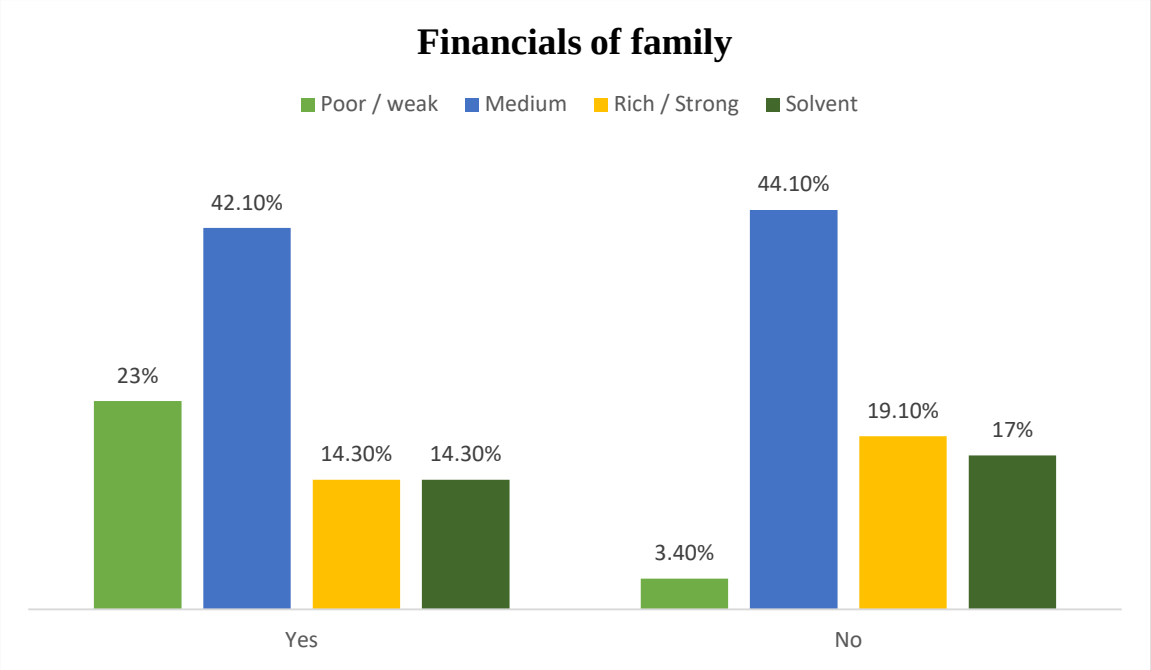


Figure 3.7: Suicidal Rate Based on Financial Conditions of Family

3.5 Proposed Methodology

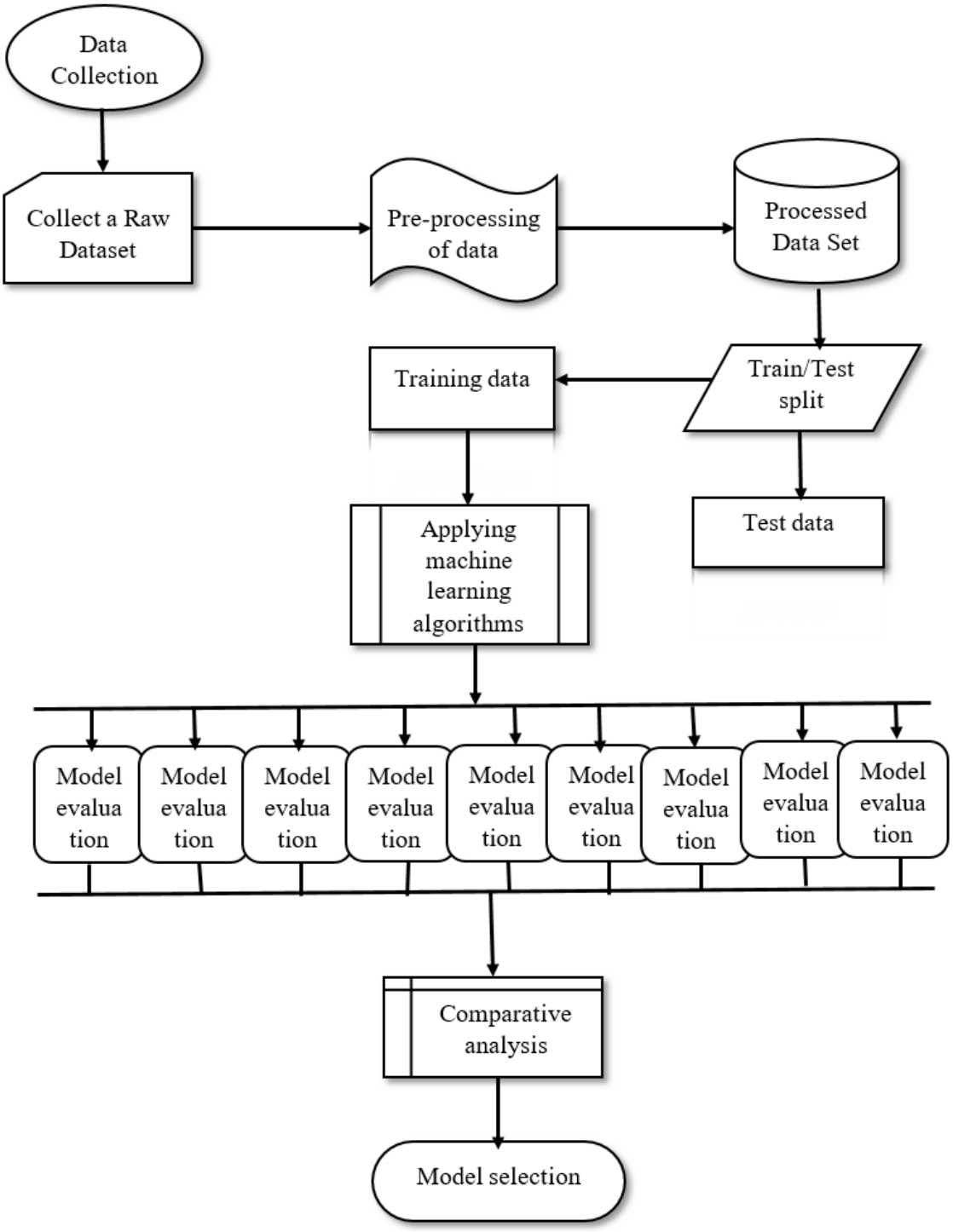


Figure 3.8: Proposed Methodology for model selection

3.6 Implementation Requirements

For implementation, first of all, we needed a Dataset that would contain all of our finding variables. After pre-processed the data, then we had need to,

- Selected of all the features
- Dropped the data duplicity
- Made certain features normalized

We collect the dataset from ‘kaggle’. After all the processes shown in the proposed methodology part.

The collected datasets saved in Microsoft Excel. The dataset contains 22 attributes and 441 instances. This study aims to build suicidal thoughts prediction models with the best accuracy. Based on collected data analysis, also have shown some charts about the percentage of people who have suicidal thoughts, how many of them have taken drug tendencies, how many people are in mental depression or loneliness shown in the statistical analysis section. For algorithm implementation, we used “Google colab”.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

The objective of the paper, dataset and how to build the model is discussed in the previous chapter. This section will carry the result of the applying algorithm. For building the model nine machine learning algorithms are used. Naive Bayes, KNN, logistic regression, support vector machine (SVM), decision tree, random forest, ADA boosting classifier and gradient boosting classifier algorithms are used. After analyzing the result, a random forest classifier achieved the highest accuracy which is 91%. After analyzing the dataset we also showed the percentage of drug-addicted people or people who have suicidal thoughts, how many of them have any smoking habit. A heatmap, scatter diagram, and histogram also shows the total calculation of these analyses.

4.2 Experimental Results & Analysis

For the suicidal prediction model, nine machine-learning algorithms are used. We compared one algorithm with another algorithm. We create all of the algorithm classification tables (include their accuracy, confusion matrix, precision, recall, F1 score). Calculate their sensitivity and specificity. Also shown, ROC curve, AUC score and confusion matrix plot.

4.2.1 Experimental Evaluation

The dataset contains 22 attributes and 441 rows. Nine machine learning algorithms are utilized to build the model. Machine learning algorithms have some hyper parameters. Using customized parameters for tuning sometimes accuracy is getting worse or sometimes it gives the best result. Hyper parameter tuning only works if a set of customized parameters make a better setup than the default setup.

There are many reasons behind calculating the accuracy of a model. Regularization, input-output data, is responsible for a model's performance. Example, Here, Input x, Output y. It's

Only classified if someone has suicidal thoughts (1) or not (2). It's called binary classification. If the classification would be multiple classification then it would be possible to check how good the model performs by changing input and output values.

Regularization is used to optimize a machine learning model. Regularization is a technique used for tuning a function by summing a penalty term in the wrong function. The added term measures the extremely varying function.

Regularization inhibits learning a more difficult or flexible model, to check to overfit. Overfitting works for better performance on the training data and makes poor performance to other data. Underfitting damages the efficiency of our machine learning model.

The total data in our dataset is 431. 20% of data has been taken for testing. 80% data is training data. Dataset has 441 rows and 22 attributes. 21 data is set as input and 1 is given for output. Nine machine learning algorithms on our dataset and got different accuracy for each. They are Naive bias, K-NN, Linear SVM, Non-linear SVM, Random Forest, Decision Tree, Logistic regression, XG-boost, Ada-boost.

First applied naive bias algorithm. Naive Bayes classifiers are a combination of classification algorithms. It is based on Bayes' Theorem. It is quick and can be used to obtain real-time predictions. It is not sensitive to unnecessary features. It sometimes gives worse data because if a definite variable has a section in the test dataset, that is not recognized in the training dataset. After that, the model will specify a 0 (zero) possibility. It will be inadequate to perform a prediction.

We got the accuracy for naive bias is 78% and its F1 score is 78%.

KNN Very simple implementation. Robust concerning the research space. Its classes don't have to be linearly divisible. It provides 82% of accuracy when the neighbor value was 2. After changing the neighbor value KNN achieved an accuracy of 87%(neighbor value = 4).

A support vector machine (SVM) is a supervised machine learning model. It employs classification algorithms for two-group classification queries. After providing an SVM model set of particularized training data for each level, they're able to classify new text. Linear and

Non-linear. After utilizing linear SVM classifiers achieved 75% accuracy. The kernel method is the real power of SVM. It compares moderately strong to high-dimensional data.

For Non-linear SVM the accuracy was 83%. The non-linear accuracy score of 100% is realistically unattainable when it comes to the real application of these techniques, and it highlights a greater potential for these types of models to overfit data.

Random forests is a supervised learning algorithm. Random forests build decision trees on randomly chosen data samples. It prepares a prediction from a separate tree and picks the best solvent utilizing deciding. It also presents a beneficial symbol of the feature's value.

Random forests are known as an extremely accurate and strong classification because of the number of decision trees combining in the method.

It does not endure the overfitting problem. The major reason is that it practices the center of all the predictions, which removes out the biases. Random forests can also manage dropping value.

We got the highest accuracy out of Random forest among all these 9 algorithms. The random forest achieved 91% accuracy which is the highest.

After training the model, the Decision tree algorithm provides 86% accuracy. Decision Tree algorithm refers to the group of supervised learning algorithms. They boost imminent models with accuracy, expertise in understanding, and security.

Logistic regression is simpler to perform and very suitable to train data. If the representation of notes is more inferior than the number of characteristics. Logistic regression may occur to overfitting. It makes no presumptions about the relationships of classes in feature season.

Logistics regress has some solvers. They are ‘newton-cg’, ‘lbfgs’, ‘liblinear’, ‘sag’, ‘saga. To find better accuracy there used three different hyper parameters. For newton-cg, the accuracy and F1-score both were 75%. Liblinear and SAG both achieved 75% accuracy.

XGBoost is a very effective algorithm. It is very easy-to-use and it passes high performance and accuracy. XGBoost is also recognized as a regularized version of GBM. XGBoost has in-built Lasso Regression (L1) and Ridge Regression (L2) regularization. It helps to restrict the model from overfitting. For that XGBoost is also named a regularized form of GBM. GMB is known as Gradient Boosting Machine. XGBoost employs the ability of parallel processing.

XGBoost has an in-built capacity to manage missing values. XGBoost enables users to operate cross-validation at each redundancy of the boosting method. XG-boost achieved an accuracy of 75% for this model.

AdaBoost is the very first boosting algorithm. It is used to be accommodated in solving methods. After implementing the Ada-boost classifier the accuracy was 72%.

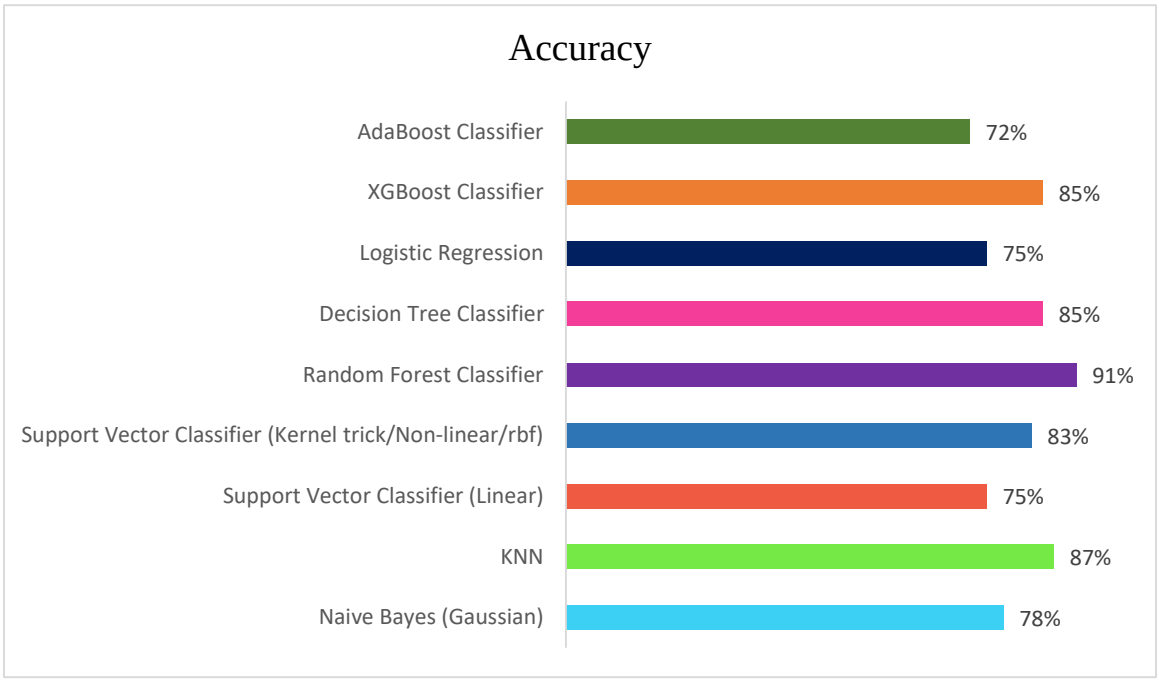


Figure 4.1: Accuracy level of applied algorithms

4.2.2 Descriptive Analysis

Supervised learning has two parts. Classification and Regression. Our model is classification-based. In classification-based algorithms, there have to check for the F1 score, precession, recall value. The only accuracy checked is not valid for the classification algorithms. To build the model we also calculated sensitivity, specificity, precision, recall, f-score, and roc-curve and confusion matrix of all the algorithms. All evolution is required for the model selection.

Classifications are estimated based on the test data set for better Analysis.

The sensitivity of a test can accommodate to determine how strong it can classify a model. Sensitivity is the true positive rate. The rules of sensitivity,

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100\%$$

Specificity is an analysis of how well a test can recognize true negatives in a test set. Specificity is calculated by,

$$\text{Specificity} = \frac{TN}{FP+TN} \times 100\%$$

Precision is the ratio of true positive value and predicted positive value. Precision is calculated by this formula,

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\%$$

Precision is the estimation of accuracy.

We also measure the F1 score, Recall value. A recall is also the analysis of accuracy. It is the proportion of true positive value and true positive value.

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\%$$

Recall is calculated by this formula.

F1 score used for both false positive and false negative values for estimation

$$\text{F1 Score} = \frac{2 \times Precision \times Recall}{Precision+Recall} \times 100\%$$

A receiver operating characteristic curve or ROC curve is a graph representing the conclusion of a classification model at an individual classification threshold. The curve plots have two parameters: One is True Positive Rate and another one is False Positive Rate. The Area under the Curve or AUC is the division of the capacity of a classifier to discriminate between classes. It is served as a review of the ROC curve.

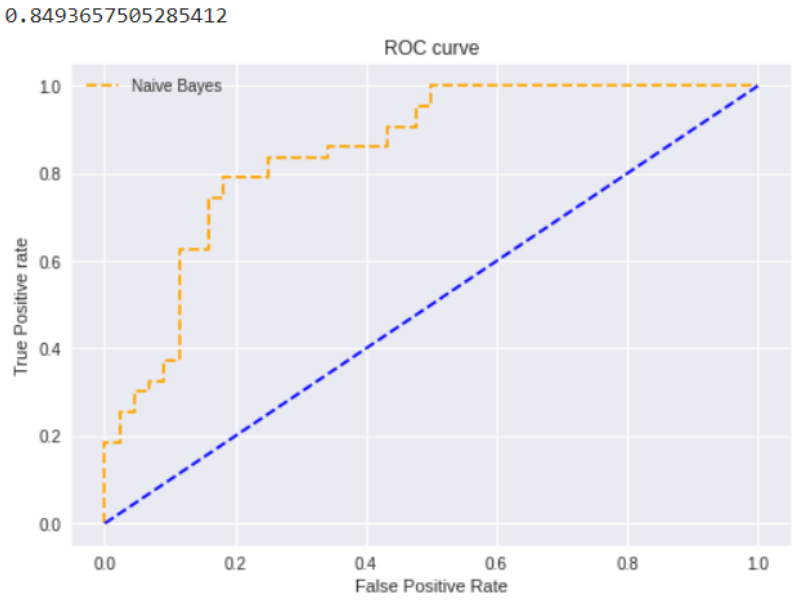


Figure 4.2: ROC curve of the Naïve Bayes (Gaussian)

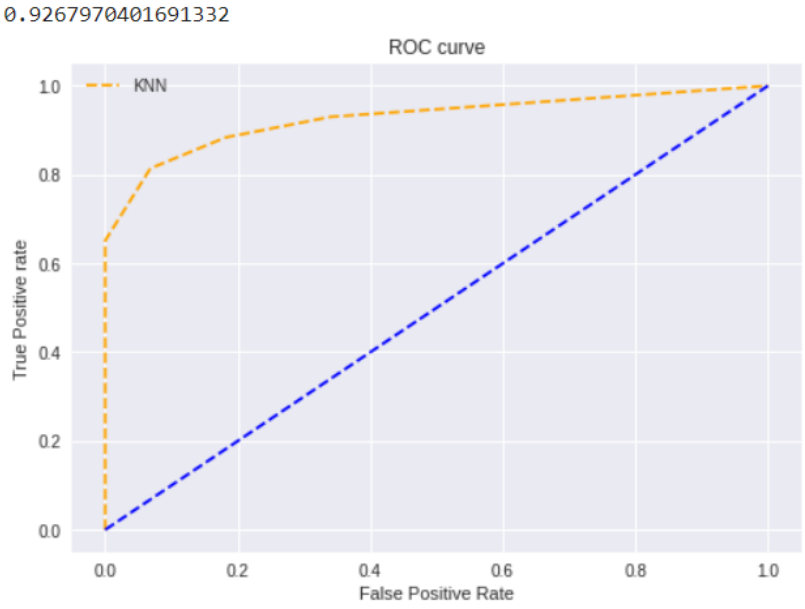


Figure 4.3: ROC curve of the KNN algorithm.

```
0.8218816067653277
```

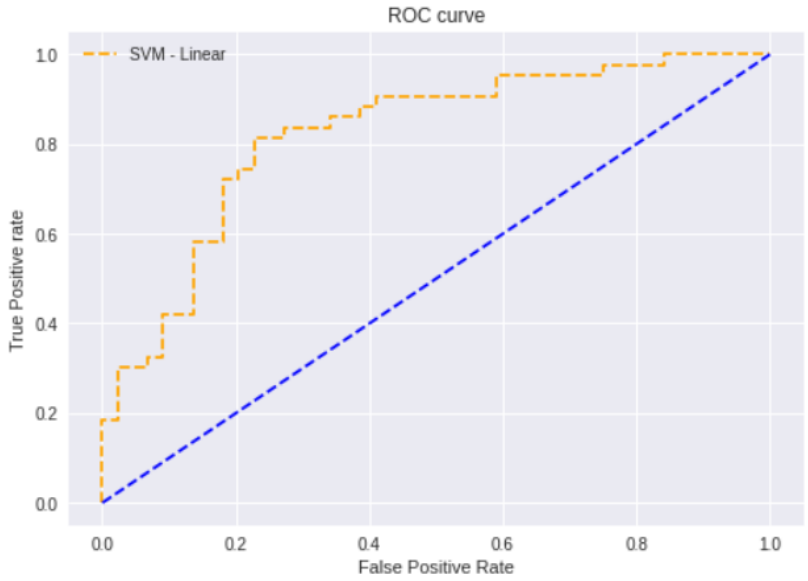


Figure 4.4: ROC curve of the Support Vector Classifier (Linear).

```
0.9408033826638478
```

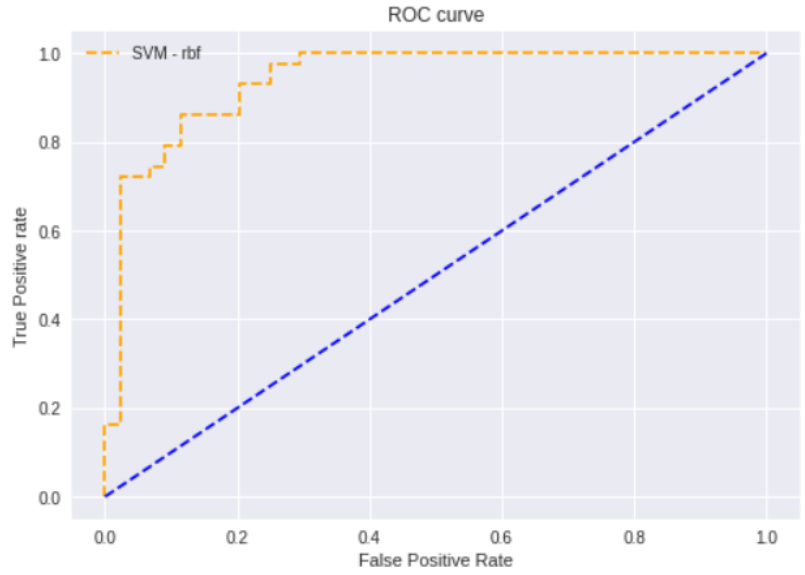


Figure 4.5: ROC curve of the Support Vector Classifier (Non-Linear)

```
0.9659090909090908
```

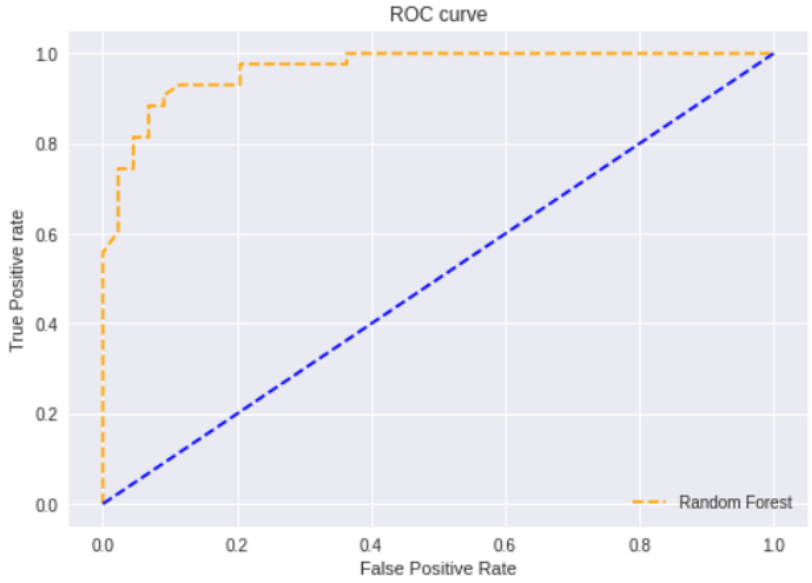


Figure 4.6: ROC curve of the Random Forest Classifier.

```
0.8506871035940803
```

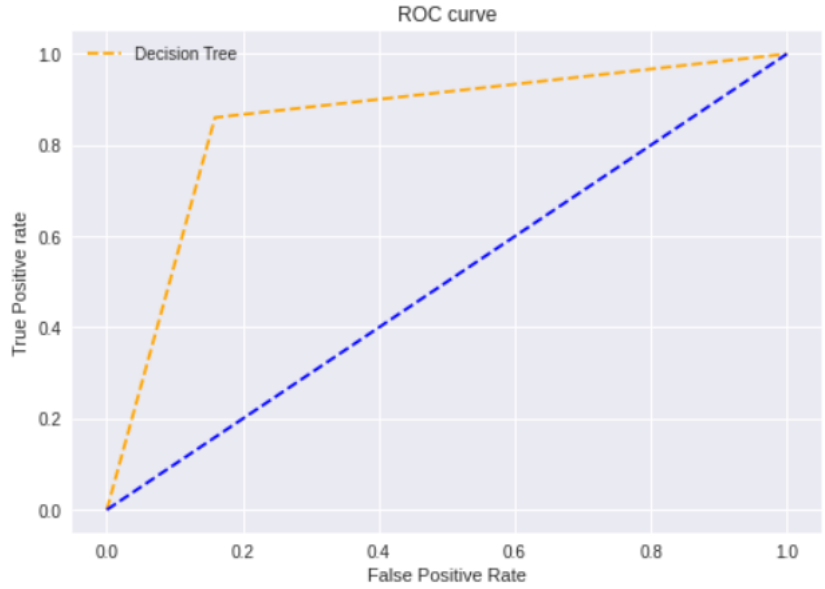


Figure 4.7: ROC curve of the Decision Tree Classifier.

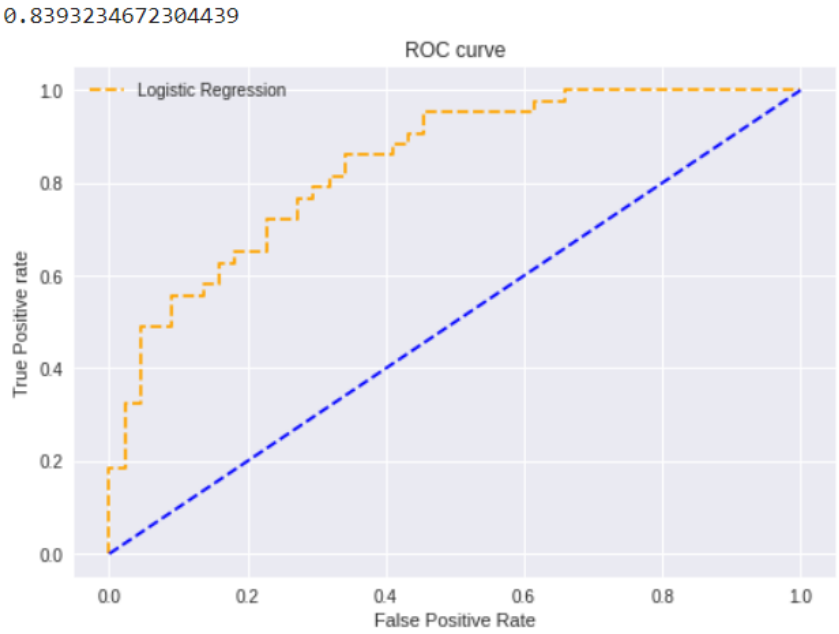


Figure 4.8: ROC curve of the Logistic Regression.

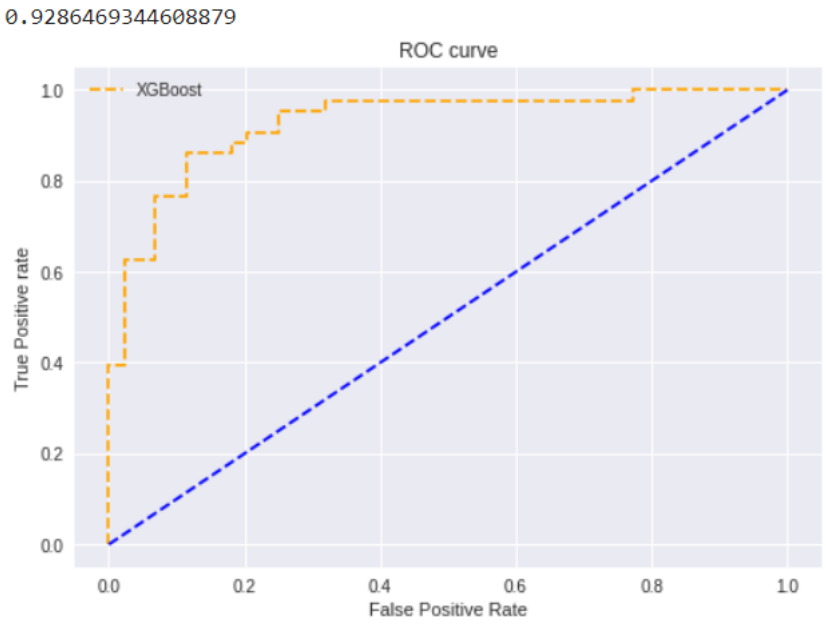


Figure 4.9: ROC curve of the XGBoost Classifier.

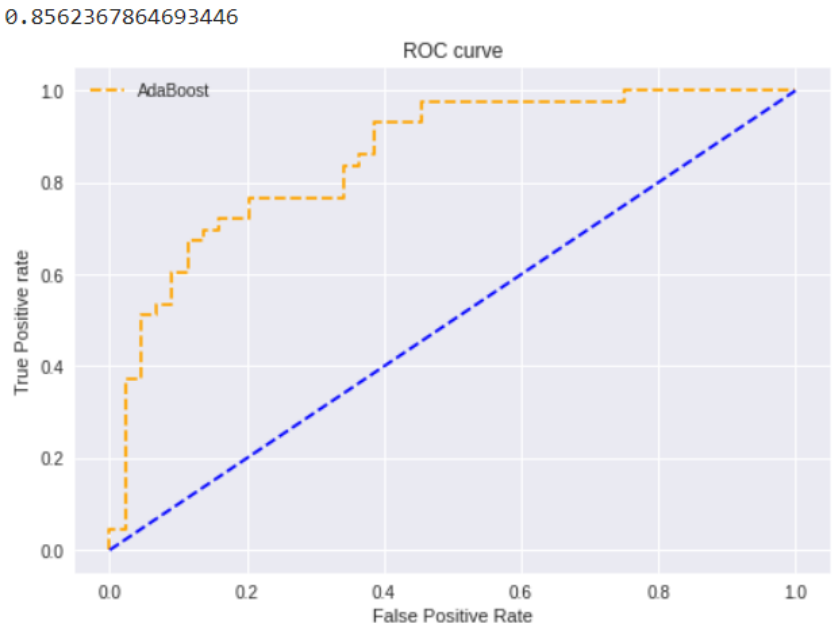


Figure 4.10: ROC curve of ADA boosting classifier.

A confusion matrix is a table that is, commonly used to justify the performance of a classification model on several test data for which the true values are classified.

Table 4.1 shows the confusion matrix of nine algorithms which are used in our model.

TABLE 4.1: CONFUSION MATRIX BASED IN CLASSIFIRES

Algorithms	Confusion matrix	Algorithms	Confusion matrix
Naive Bayes (Gaussian)	True class / No: 32, 12 / Yes: 7, 36 / Predicted label (No, Yes)	K-Nearest Neighbors	True class / No: 41, 3 / Yes: 8, 35 / Predicted label (No, Yes)
Support Vector Machine (Linear)	True class / No: 29, 15 / Yes: 7, 36 / Predicted label (No, Yes)	Support Vector Machine (Kernel trick/Non-linear)	True class / No: 35, 9 / Yes: 6, 37 / Predicted label (No, Yes)

Random Forest Classifier	True class		No	Yes	Decision Tree Classifier	True class		No	Yes
		No	40	4			No	38	6
		Yes	4	39			Yes	8	35
		Predicted label					Predicted label		
Logistic Regression	True class		No	Yes	XG Boost Classifier	True class		No	Yes
		No	30	14			No	37	7
		Yes	8	35			Yes	6	37
		Predicted label					Predicted label		
Ada Boost Classifier	True class		No	Yes					
		No	30	14					
		Yes	10	33					
		Predicted label							

Table (4.2) describes the performance of all algorithms. Measuring all the accuracy of each algorithm Random forest contains best accuracy. Again F1 score (91%), precession (91%), recall (91%), sensitivity (91%), Random forest performs the best. KNN=4 contains the high Specificity (92%).

TABLE 4.2: PERFORMANCE OF CLASSIFIRES

Algorithms		Accuracy	Sensitivity	Specificity	Precision	Recall	F1-Score
NB (Gaussian)		78%	78%	75%	79%	78%	78%
K-Nearest Neighbors	(neighbors value=2)	83%	80%	90%	83%	80%	83%
	(Neighbors value=4)	87%	87%	92%	88%	87%	87%
SVM (Linear)		75%	75%	70.5%	76%	75%	75%

SVM (Nonn-linear)		83%	83%	80.4%	83%	83%	83%
RFC		91%	91%	90.7%	91%	91%	91%
DT		85%	85%	35.4%	85%	85%	85%
Logistic Regression	(solver=newton-cg)	75%	75%	71.4%	75%	75%	75%
	(solver=saga)	75%	75%	71.4%	75%	75%	75%
	(solver=liblinear)	75%	75%	71.4%	75%	75%	75%
XGBoost Classifier		85%	85%	84.1%	85%	85%	85%
AdaBoost Classifier		72%	72%	70.2%	73%	72%	72%

4.2.3 Comparative Analysis

Some research papers worked on suicidal attempts, suicidal ideation, suicidal risk prediction models. Paper [1] predicted suicidal attempts using random forest classifiers, achieving the accuracy 83.7%. Another study [2] predict suicidal ideation and using random forest classifier they got the accuracy 82.1%. Paper [3] predicted suicidal ideation and used a conventional logistic regression model and achieved the high accuracy is 86.7%. Lastly, research paper [4] did a prediction model with extreme gradient boosting algorithm and get 79% accuracy.

This paper predicted suicidal thoughts, models based on machine learning algorithms and using the Random forest classifier we got the highest accuracy (91%) which is the heights among all of the reference paper.

TABLE 4.3: COMPARATIVE ANALYSIS

Reference	Subject	Problem Domain	Sample size	Algorithm	Accuracy
This paper	Suicidal Thought	Prediction	441	Random Forest Classifier	91%
[27]	Suicidal Attempt	Prediction	469	Random Forest Classifier	83.7%
[28]	Suicide Ideation	Prediction	11,628	Random Forest Model	82.1%
[18]	Suicidal Attempts	Prediction	7306 abstracts reviewed	Global classification	80%
[11]	Suicidal Ideation	Prediction	Training dataset 16,437, Testing dataset 3,788	Conventional logistic regression	86.7%

4.3 Discussion

This section reviews the performance of all nine machine learning algorithms accuracy, sensitivity, specificity, recall, precision, F1 score, and ROC curve. The equations of developed models and their functions are also described here. After analyzing, the random forest classifier algorithm achieved the highest accuracy with 91%. The RF classifier achieved 91% sensitivity, 91% specificity, 90.24% precision, 91% recall, and 91% F1 score. The suicidal thoughts prediction model performs better with the Random Forest classifier.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Family

In life situations, we all believe that there must be a solution, even if we are facing difficulties. Some unhealthy lifestyle activities hinder people from getting into the depths of depression and as a solution they commit suicide to relieve themselves. Since suicide is an act that ends your life, the consequences for family members are often fatal. According to research, young people between the ages of 15 and 24 prefer this inexplicable option, but the older group is less than 40 years old. This loss can leave a lasting impression on parents and friends. It affects family communication and if there are minor members, the development process is interrupted. In some cases, the victim may be the only person receiving the family. As a result of suicide, a financial crisis arises and the whole family collapses. This loss sometimes creates suicidal tendencies in others when nothing is left. There is conflict of mind, painful depression and grief. Since our approach can reverse suicidal tendencies, using our machine-based assessment tools can prevent suicide attempts and protect its members so that the family can claim to prevent death if given a chance in the future. When abnormal activity is detected, the model takes several approaches, such as input and whether or not the effect we are getting is a trend. If existence is clear, caregivers should be deeply concerned, honest and attentive to their children's work. These days children suffer from severe depression at an early age. While the model may have a positive effect, parents should consider the cause of the depression. Our model states that the same happens where there is a shortage of family ties. Prejudice can have a devastating effect on children who, with proper guidance and information, can overcome in our path when they feel abandoned by their loved ones. With this model, patterns of suicide are identified and suicides are not lost on others. Therefore, a family can monitor their children and prevent unhealthy activities with awareness and honesty.

5.2 Impact on Society

To live, we need to attach to society because people by nature are social creatures. We need a sense of belongings and that sense of belonging is what connects us to many relationships we

develop and we surrounded by people who share similar values. So, if there is any loss it directly impacts society. Where the matter of suicide, people of society affected in various ways. This term can be followed by others in society, as we take both positive and negative learning from communication. A frightening and alarming situation occurs, the normal life of people is disrupted and communication of people in society becomes immobilize. As a result, social disorder and anarchy are created. This activity decreases the status of society and is the ignominy of society. To prevent this situation our model of suicide tendency prediction can be used. The reason of thoughts which compel them to commit suicide, we can assume and by consultation, this thought can be removed and guide them also in a right way. The aged people mostly suffer from depression of financial status. Sometimes, they may not satisfied with their workplace. The responsibility of bearing the whole family becomes impossible to carry out. This certain failure is the reason behind a suicide that can be detected through the model by their activity. By assuming the facts, people can stand by each other. People in society need to be responsible for each other and come forward in danger. The proposed approach will give results depending on the suspect's information by which people in the society can take appropriate steps to prevent the attempt. Analyzing the data, they'll be able to find out the reason, and after knowing that consultation should be done. Thus, with the help of the prediction model social order, fame, connectivity of people in society will remain stable and strong. This observation will lead society to the advancement and will elevate the status of society.

5.2.1 Impact on Attempt-Survivor

Families and families surviving such attempts, including suicide attempts, suffer a lot of damage. They suffer from humiliation for their attempt. Mental turmoil makes them more apprehensive. In such a situation, it is not possible for them to express the cause of mental depression where the procedure of consultation depends on the reason. As we have brought the technological benefit through our model, the reason can be identified. The motive of mental breakdown can be found out on the basis of some general information, which will not actually put pressure on their mental state. And they can be easily guided to the right path. Relying on analyzed data, people can aware of all those motives which could be influenced them again. Incidents that can put pressure on their mental state can be avoided. And all these steps any

people can take through the information analyzed by our model. Above all, this prediction approach will play an important role in bringing attempted survivors to the path of light and healing them.

5.3 Moral Aspects

People do not provide their sensitive information without too much credibility. An experiment or analyzing approach should be on all the information that does not interfere with personal information. Keeping that in mind, we have established our model and worked with all that information which will not violate human rights and will be no chance of immorality. Here some examples of the information that we worked on-age, gender, education, having mental issues or not, financial state, family relations, etc. Collecting all those information will not intervene with any individual's affairs, rather it will keep their family and society informed of any risk. All types of collected information will be kept confidential, their privacy will be conserved and by emphasizing on the concealed matter, our introduced model will continue its task to achieve the purpose.

5.4 Sustainability Plan

The sustainability plan shows, many perspectives of documentation where the approach has to be maintained in a sustained way to still function. This document focuses on social stability, financial stability and organizational stability. Our quest is to prevent suicidal attempts through analyzing the behavior and activities which is able to find out the tendency of suicide. The model is organized in such a way that, people of all classes can easily operate and get benefited. It is designed to ensure its use and sustainability socially. Various consultation centers, psychiatrists, suicide prevention agencies and government affiliates organizations will be able to use this model by which society and nation will be defended from possible harm. Thus the model will ensure permanency through the development of society.

CHAPTER 6

SUMMARY, CONCLUSION AND IMPLICATION FOR FUTURE

6.1 Summary of the Study

For implementing nine machine learning algorithms, we need to follow some steps:

Step 1: Dataset collection.

Step 2: The collected data make summarized.

Step 3: Data preprocessing.

- 3.1: Duplicity dropping.
- 3.2: Null Value Checking.
- 3.3: SMOTE Technique.
- 3.4: Feature Selection.
- 3.5: Data Normalization.

Step 5: Classification analysis.

- 5.1: Train/Test split.
- 5.2: Dimensionality Reduction technique.
- 5.3: Training Model: Logistic regression, Decision Tree classifier, Random forest classifier.

Step 6: Evaluate the Model.

Step 7: Model Selection.

Step 8: Parameter Tuning.

Step 9: Make Predictions.

Step 10: Results.

- 10.1: Comparative Result/Best result.
- 10.2: Confusion matrix.
- 10.3: Classification report.

Step 11: Analysis.

Step 12: Check the final result.

6.2 Limitations and Conclusion

This model works for a limited sequence the dataset is not much enough for implementation. Due to lockdown and COVID-19, we could not go outside for data collection. So there is a scope to work with a huge dataset. Many advanced methods could also be used for data processing, and the model could be presented beautifully using different variations in the application of algorithms. It is possible to get better accuracy by utilizing more different algorithms. Suicidal thought has become a vital issue in the current scenario where people are doing suicide nowadays on various causes. This increasing trend of suicide need to turn into downward trend. For that reason, we have developed a model that is able to predict if a person is in a mindset of doing suicide or not. A comparative analysis among different classifier techniques is done to find out the most optimal technique to identify the suicidal thought of a person. In this thesis, this approach was used to predict the suicidal mindset of a person using supervised classifier technique. Among all, we found "Random Forest Classifier" technique as the most reliable and efficient method out of 9 classifier technique with an accuracy level of 91%. For a sustainable environment, this model will help to mitigate the consequences of future development of people and their living nature to cope up with the environment.

6.3 Implication for Further Study

This prediction model will be more improved in the future. Research work requires implementation or development for future implementation. Researchers are looking for other research limitations, and their future implementation is reliant on the limitations of the other earlier research work. Our model is a machine learning based prediction model. We showed a model with a better accuracy algorithm. There is a possibility to make this model on android application or web application. In future it is possible to increase the accuracy of our model using a larger database by creating user-friendly GUI or using any software development. Most of the implementation is just academic, if proper funding is available, then we will be able to make this model available to everybody so that anyone can check if someone has any suicidal thoughts.

REFERENCES

[1] Suicide. (2019). Retrieved 9 May 2021, from https://www.who.int/news-room/fact-sheets/detail/suicide

[2] Suicide in Bangladesh - Wikipedia. (2021). Retrieved 9 May 2021, from https://en.wikipedia.org/wiki/Suicide_in_Bangladesh

[3] Bangladesh: Suicide claims more lives than coronavirus. (2021). Retrieved 9 May 2021, from https://www.aa.com.tr/en/asia-pacific/bangladesh-suicide-claims-more-lives-than-coronavirus/2175200

[4] (2021). Retrieved 9 May 2021, from https://apps.who.int/iris/bitstream/handle/10665/254982/9789290225737-eng.pdf

[5] Examining the alarming suicide trends in Bangladesh. (2018). Retrieved 9 May 2021, from https://www.dhakatribune.com/opinion/special/2018/05/08/examining-alarming-suicide-trends-bangladesh

[6] Social alienation triggers suicidal tendencies among adolescents. (2021). Retrieved 8 May 2021, from https://www.unicef.org/bangladesh/en/stories/social-alienation-triggers-suicidal-tendencies-among-adolescents

[7] Suicide in Children and Teens . (2021). Retrieved 8 May 2021, from https://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/FFF-Guide/Teen-Suicide-010.aspx

[8] Birjali, Marouane & beni hssane, Abderrahim & Erritali, Mohammed. (2017). Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks. Procedia Computer Science. 113. 65-72. 10.1016/j.procs.2017.08.290.

[9] Hettige, Nuwan C., et al. "Classification of suicide attempters in schizophrenia using sociocultural and clinical features: A machine learning approach." General hospital psychiatry 47 (2017): 20-28.

[10] Jung, Jun & Park, Sung & Kim, Eun & Na, Kyoung-Sae & Kim, Young Jae & Kim, Kwanggi. (2019). Prediction models for high risk of suicide in Korean adolescents using machine learning techniques. PloS one. 14. e0217639. 10.1371/journal.pone.0217639.

[11] Ryu, Seunghyong & Lee, Hyeongrae & Lee, Dong-Kyun & Park, Kyeongwoo. (2018). Use of a Machine Learning Algorithm to Predict Individuals with Suicide Ideation in the General Population. Psychiatry Investigation. 15. 10.30773/pi.2018.08.27.

[12] Roy, Arunima & Nikolitch, Katerina & McGinn, Rachel & Jinah, Safiya & Klement, William & Kaminsky, Zachary. (2020). A machine learning approach predicts future risk to suicidal ideation from social media data. npj Digital Medicine. 3. 10.1038/s41746-020-0287-6.

[13] Mens, Kasper & Schepper, CWM & Wijnen, Ben & Koldijk, Saskia & Schnack, Hugo & de Looff, Peter & Lokkerbol, Joran & Wetherall, Karen & Cleare, Seonaid & O'Connor, Rory & De Beurs, Derek. (2020). Predicting future suicidal behaviour in young adults, with different machine learning techniques: A population-based longitudinal study. Journal of Affective Disorders. 271. 10.1016/j.jad.2020.03.081.

[14] Ballard, Elizabeth & Voort, Jennifer & Luckenbaugh, David & Machado-Vieira, Rodrigo & Tohen, Mauricio & Zarate, Carlos. (2016). Acute risk factors for suicide attempts and death: Prospective findings from the STEP-BD study. Bipolar Disorders. 18. 10.1111/bdi.12397.

[15] Belsher, Bradley & Smolenski, Derek & Pruitt, Larry & Bush, Nigel & Beech, Erin & Workman, Don & Morgan, Rebecca & Evatt, Daniel & Tucker, Jennifer & Skopp, Nancy. (2019). Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. JAMA Psychiatry. 10.1001/jamapsychiatry.2019.0174.

[16] Lyu, Juncheng & Zhang, Jie. (2018). BP Neural Network Prediction Model for Suicide Attempt among Chinese Rural Residents. Journal of Affective Disorders. 246. 10.1016/j.jad.2018.12.111.

[17] Barak-Corren, Yuval & Castro, Victor & Javitt, Solomon & Hoffnagle, Alison & Dai, Yael & Perlis, Roy & Nock, Matthew & Smoller, Jordan & Reis, Ben. (2016). Predicting Suicidal Behavior From Longitudinal Electronic Health Records. The American journal of psychiatry. 174. appiajp201616010077. 10.1176/appi.ajp.2016.16010077.

[18] Belsher, Bradley & Smolenski, Derek & Pruitt, Larry & Bush, Nigel & Beech, Erin & Workman, Don & Morgan, Rebecca & Evatt, Daniel & Tucker, Jennifer & Skopp, Nancy. (2019). Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. JAMA Psychiatry. 10.1001/jamapsychiatry.2019.0174.

[19] Pinto, Sara & Soares, Joana & Silva, Alzira & Curral, Rosário & Coelho, Rui. (2020). COVID-19 Suicide Survivors—A Hidden Grieving Population. Frontiers in Psychiatry. 11. 10.3389/fpsyt.2020.626807.

[20] Sakib, Najmuj & Islam, Merajul & Habib, Md & Bhuiyan, A K M & Alam, Md & Tasneem, Noshin & Hossain, Moazzem & Shariful Islam, Sheikh Mohammed & Griffiths, Mark & Mamun, Mohammed. (2020). Depression and suicidality among Bangladeshi students: Subject selection reasons and learning environment as potential risk factors. Perspectives In Psychiatric Care. 10.1111/ppc.12670.

[21] Pervin, Mst & Ferdowshi, Nafiza. (2016). Suicidal ideation in relation to depression, loneliness and hopelessness among university students. Dhaka University Journal of Biological Sciences. 25. 57. 10.3329/dujbs.v25i1.28495.

[22] Rasheduzzaman, M. & al Mamun, Firoj & Faruk, Md & Hosen, Ismail & Mamun, Mohammed. (2021). Depression in Bangladeshi university students: The role of socio-demographic, personal and familial psychopathological factors. Perspectives In Psychiatric Care. 10.1111/ppc.12722.

[23] Khan, Md. Mostaured Ali & Rahman, Mosfequr & Islam, Md & Karim, Masud & Hasan, Mahmudul & Jesmin, Syeda. (2020). Suicidal behavior among school-going adolescents in Bangladesh: findings of the global school-based student health survey. Social Psychiatry and Psychiatric Epidemiology. 55. 10.1007/s00127-020-01867-z.

[24] Mamun, Mohammed & Akter, Tahmina & Zohra, Fatema & Sakib, Najmuj & Bhuiyan, A K M & Banik, Palash & Muhit, Mohammad. (2020). Prevalence and risk factors of COVID-19 suicidal behavior in Bangladeshi population: are healthcare professionals at greater risk?. Heliyon. 6. 10.1016/j.heliyon.2020.e05259.

[25] Choudhury, Ahnaf & Khan, Md Rezwan & Nahim, Nabuat & Tulon, Sadid & Islam, Samiul & Chakrabarty, Amitabha. (2019). Predicting Depression in Bangladeshi Undergraduates using Machine Learning. 789-794. 10.1109/TENSYMP46218.2019.8971369.

[26] Hasan, M.Tasdik. (2020). Depression, sleeping pattern & suicidal ideation among medical students in Bangladesh: A cross-sectional pilot study. 10.31234/osf.io/qk24v.

[27] Hasan, Tanvir & Hawlader, Muhammad Rajib & Salekin Khan, Md. Serajus. (2020). Towards Developing a Machine Learning Model For Suicidal Attempt Prediction.

[28] Oh, Bumjo & Yun, Je-Yeon & Yeo, Eun & Kim, Dong-Hoi & Kim, Jin & Cho, Bum-Joo. (2020). Prediction of Suicidal Ideation among Korean Adults Using Machine Learning: A Cross-Sectional Study. Psychiatry Investigation. 17. 10.30773/pi.2019.0270.

PLAGIARISM REPORT

ORIGINALITY REPORT

17%	12%	7%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	2%
3	Submitted to Columbia High School Student Paper	1%
4	Submitted to Mansoura University Student Paper	1%
5	dspace.library.daffodilvarsity.edu.bd:8080 Internet Source	1%
6	Sara Pinto, Joana Soares, Alzira Silva, Rosário Curral, Rui Coelho. "COVID-19 Suicide Survivors—A Hidden Grieving Population", Frontiers in Psychiatry, 2020 Publication	<1%
7	Marouane Birjali, Abderrahim Beni-Hssane, Mohammed Erritali. "Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in	<1%