Suicide Attempt Prediction Rate Analysis Based on the Machine Learning Approach

Riad Al-Ashekin

ID: 2015–2–60–034

Saqlain Mustak Sagor

ID: 2016-3-60-015

Muhammad Farabi Hossain

ID: 2016–3–60–051

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering

Bachelor of Science in

Computer Science and Engineering

Department of Computer Science and Engineering

East West University

Dhaka-1212, Bangladesh

June, 2021

ABSTRACT

Suicide is a very critical issue in this world. If we can detect early in this, then we can prevent it and can save a life. Every day we can see in our newspaper many peoples commit suicide in this world. Now we can see Suicide is one of the major causes of death day by day in this world. Many people attempt suicide in this world. We can see the basic reason for many people is depression. Every year around 780,000+ annual deaths up to 20+ million suicide attempts. In the United States, suicide rates increased every year. Which is around 45,000+ per year. Suicide death rates are increasing day by day in any country. In our nearest country India, the Indian suicide rate is 16.4 per 100,000 in women and 25.8 for men. In our nearest country, Pakistan suicide rate is 7.28 per 100,000. In Bangladesh, suicide is a very critical issue. Every year Bangladeshi peoples dead in suicide around 2.06% worldwide. In the police and hospital report, every year increasing the death rate. We can see the suicide commit rate in Bangladeshi people which is 128.08 per 100,000. In this Artificial Intelligence system, we are training a machine to train our own data set. we are using in our collecting data set which are collected by our friends and family members. They all are living in Bangladesh. We are collecting this data some frequently ask questions like gender, age, income, body weight, friends, depression, attempt suicide, employment, job title, education level, improve yourself, etc. In this system, we are using four algorithms for comparison and prediction purposes which algorithms name are the Random forest algorithm, XG Boost algorithm, Support Vector Machine (SVM) algorithm, Light GBM algorithm. In this system, we use another data set which is outside Bangladesh. In this data set, we use for comparison purposes for our country data set. This Artificial intelligence system can help us early detect and prevent suicide attempts people. It can also help medical technologists, doctors, psychiatrists, social workers, and any other people.

DECLARATION

We declare that this thesis and the dataset, method which we are using in this system, this dataset is our own and collected by our team members. We use our developed method in this system. We hereby declare that this thesis title **"Suicide Attempt Prediction Rate Analysis Based on the Machine Learning Approach"** which we are submitted in fulfillment of the requirements for the degree of Bachelor of Science in computer science and engineering in the faculty of computer science and engineering, East West University. This is our work. We are showing the result of our work. We declare that to the best of our knowledge, it has no part of this report has been submitted elsewhere for any other degree or diploma or professional qualification, except where due reference is made in the text of this thesis.

Supervisor: .....

Dr. Maheen Islam

Associate Professor, Department of CSE,

East West University.

Riad Al-Ashekin.

ID: 2015-2-60-034: .....

Saqlain Mustak Sagor

ID: 2016-3-60-015: .....

Muhammad Farabi Hossian

ID: 2016-3-60-051: .....

ACKNOWLEDGEMENT

At first, we wish to convey our cordial thanks and gratitude to almighty God for everything. We would like to thanks our parents and everyone who has supported us all the way through to complete this thesis successfully. We would like to express our gratitude to our supervisor **Dr. Maheen Islam,** Associate Professor, Department of computer science and engineering, East West University, for her guidance, inspiration, valuable suggestion, encouragement, thorough review and correction of some working mistake, without her we cannot finalize in this system. We feel very proud to work with her. Thanks to our computer science and engineering department chairman and all department chairman and faculty members. We would like to thank all staff and graduate students. We wish to express our gratitude to East West University for providing us an excellent research environment and all other facilities to complete this thesis.

With Best Regards,

Riad Al-Ashekin.

ID: 2015-2-60-034: .....

Saqlain Mustak Sagor

ID: 2016-3-60-015: .....

Muhammad Farabi Hossian

ID: 2016-3-60-051: .....

APPROVAL

This thesis title **"Suicide Attempt Prediction Rate Analysis Based on the Machine Learning Approach"** Submitted by Riad Al-Ashekin ID: 2015-2-60-034, Saqlain Mustak Sagor ID: 2016-3-60-015, Muhammad Farabi Hossian ID: 2016-3-60-051, Department of Computer Science and Engineering (CSE), East West University (EWU) under the supervision of Dr. Maheen Islam, Associate Professor, Department of Computer Science and Engineering he has been accepted as satisfactory for the fulfillment of the requirements for the degree of Bachelor of Science BSc. in Computer Science and Engineering and approved as to its style and contents.

Supervisor: .....

Dr. Maheen Islam

Associate Professor, Department of CSE,

East West University.

Chairman: .....

Dr. Taskeed Jabid

Chairperson and Associate Professor, Department of CSE,

East West University.

Copyright

Table of Contents

List of Table

List of Figure

Chapter-1

Introduction

1.1 Introduction:

Suicide is a major problem in today's world. Suicide is the world's leading cause of death. If we can detect this early on, we can prevent it and save a life. Every day, we read in the news about how many people commit suicide around the world. Suicide is becoming one of the leading causes of death in the world. In this world, many people attempt suicide. According to the World Health Organization's yearly report on suicide data, suicide has now become a very serious public health issue (WHO). We can see the root cause of many people's depression. According to a WHO report, approximately 800,000 people die by suicide each year, with an additional 20 million attempting to commit suicide. Suicide rates in the United States have risen year after year. This equates to approximately 45,000 people per year [1]. Suicide death rates are increasing in every country. The reasons that people commit suicide are complicated. People with depression are highly likely to commit suicide, but many without depression can also have suicidal thoughts [2]. Eighty percent of suicides occur in middle- and low-income countries. We can see that in our neighboring country, India, the female suicide rate is 16.4 per 100,000, while the male suicide rate is 25.8. Pakistan has a suicide rate of 7.28 per 100,000 people. Suicide is a major problem in Bangladesh. Every year, approximately 2.06 percent of Bangladeshi people die by suicide worldwide. According to police and hospital reports, the death rate is increasing year after year. We can see that the suicide rate in Bangladesh is 128.08 per 100,000 people. Our goal is to gather data and identify risk factors that explain and predict suicidal behavior. We can see that the predictive ability of the suicide research field has not improved in the last 49 years. We are attempting to improve our traditional prediction techniques in order to use the Machine Learning prediction technique. The Machine Learning suicide prediction technique is more effective than the traditional technique because it is more effective in significant relationships, such as using different predictors in simple regression models, which are insufficient to model multifactorial complex behavior such as suicide. We are training a machine to train our own data set in this Artificial Intelligence system. We are using data sets collected by our friends and family members in our data collection. They are all from Bangladesh. We are gathering this data by asking frequently asked questions such as gender, age, income, body weight, friends, depression, suicide attempt, employment, job title, education level, improve yourself, and so on. In this system, four algorithms are used for comparison and prediction: the Random Forest algorithm, the XG Boost algorithm, the Support Vector Machine (SVM) algorithm, and the Light GBM algorithm. We use another data set that is not from Bangladesh in this system. We use this data set for comparison

purposes with our country data set. This artificial intelligence system can assist us in detecting and preventing suicide attempts. It can also be beneficial to medical technologists, doctors, psychiatrists, social workers, and others.

1.2 Aim:

Our aims are to collect data and identify risk factors for people who attempt suicide. Suicidal behavior in Bangladeshi people is explained and predicted. If we can detect this slightly earlier on, we can prevent it and save a life. We are using our proposed system to identify people who are at risk of attempting suicide. We are implementing our proposed framework in our dataset collection. The results of our proposed system are very impressive.

1.3 Motivation:

Suicide is a major problem in today's world. Suicide is the world's largest cause of death.
If we can detect this early on, we can prevent it and save a life. We find that several suicide attempt prediction systems are being implemented in various countries. We all know that people in each country are unique. We can see that classic prediction methods are more challenging and have lower accuracy. We are attempting to improve our old prediction approaches in order to use the Machine Learning prediction methodology. The machine learning suicide prediction algorithm outperforms the old method. Many suicide attempt prediction systems can be found in artificial intelligence systems. This is an algorithm for predicting people's suicide attempt. We can find numerous techniques that name regular logistic regression, nearest neights to K, classification tree, random forests, vector machine gradient enhancement and supports, advanced machine learning algorithms and profound neural networks. We consider suicide in Bangladesh as a really significant problem. In Bangladesh there is no suicide prediction system.
That is why we are attempting to collect data about our country's population and apply some algorithm to assess that data. Also, compare data from other countries. We are attempting to collect data in order to uncover risk factors that explain and predict suicidal behavior. We are teaching a machine to train our own data set in our Artificial Intelligence system. We are leveraging data sets obtained by our friends and family members in our data collection. They are all from Bangladesh. We are gathering this data by asking often asked questions such as gender, age, income, body weight, friends, depression, suicide attempt, employment, work title, education level, improve yourself, and so on.
In this system, four algorithms are used for comparison and prediction: the Random forest methodology, the XG Boost method, the Support Vector Machine (SVM) methodology, and the Light GBM method. We use another data collection that is not from Bangladesh in this system. We utilize this data set for comparison reasons with our national data set. This artificial intelligence technology can assist us in detecting and preventing suicide attempts. It can also be beneficial to medical technicians, doctors, psychiatrists, social workers, and others.

1.4 Contributions:

To overall contributions of our research

- We introduce a suicide attempt prediction system that’s identified Suicide attempt people.
- We are applying some new methodology and techniques to analyze and Comparison to other country data.
- We experiment with two data set for comparison purposes. The first One is our Bangladeshi people dataset Second one is the other country Dataset.
- We evaluate our proposed methodology performance in suicide attempt prediction.

1.5 Organization of this research:

The organization of this thesis work follows:

- Chapter-2 in this chapter we are discussing the existing system and literature review. We are discussing many types of suicide attempt prediction system and their methodology and their limitations.
- Chapter-3 in this chapter we are discussing our system methodology, also discuss different types of algorithms that we are used in this system. Here details to discuss our overall procedure.
- Chapter-4 in this chapter we are discussing comparison and result analysis of the system results.
- Chapter-5 in this Chapter we are discussing our system affection in our society, help of our society. Here shows our system Prediction Result Accuracy Analysis.
- Chapter-6 in this chapter we conclude our overall system.

1.6 Summary:

In this chapter, we are describing our system overview. This chapter in our introductory chapter. In this section, we show our overall system overview step by step, as an introduction, as well as some related systems, motivations, objectives, and our contributions to this system. This chapter discusses our overall system procedure and how to solve our specific problem.

Chapter-2

Existing System / Literature Reviews

2.1 Introduction:

We'll talk about existing systems or literature reviews in this chapter. Suicide is a major problem in the globe today. Suicide is the world's biggest cause of death. We can prevent it and save a life if we can recognize it early enough. Many researchers are attempting to develop methods that can detect persons who are attempting suicide. We've included some examples of suicide attempt prediction systems as well as their drawbacks. How to use the existing system, their forecast approach, and model, as well as where to put their system in place, all focus on the existing system's accuracy.

2.2 Existing System:

The related works and the literature review detailed below.

2.2.1 Prediction of attempted suicide in men and women with crack-cocaine use disorder in Brazil:

This technique was created for cocaine addicts in Brazil. They were told how to forecast how a cocaine user would try suicide and why they would do so. This research was conducted in a cross-sectional manner. Using a database that has been compiled as well. This database contains information on patients who have been hospitalized for cocaine use disorder and their clinical reports. After viewing this suicide attempt, 34 percent of males and 50 percent of women attempted suicide. They discovered that hallucinations, despair, and early trauma are all linked to suicide attempts in both men and women. In this study, we discovered that the majority of drug users attempt suicide or suffer from a mental illness. Data curation and GLM were implemented using IBM SPSS syntax, the KDD process, R scripts written in Studio IDE, and the packages VIM, mouse, DMwR, Caret, random Forest, and pROC for curves. [3]

2.2.2. Predicting future suicidal behavior in young adults, with different machine learning techniques: A population-based longitudinal study:

Using a comparison of the following techniques, this system sought to survey the output of machine learning to predict posterior suicide behavior based on population-based longitudinal data. They were gathering information on 3508 young adults (18-34 years) in Scotland who had completed a psychiatric questionnaire, as well as information on persons who had attempted suicide. One year later, they were followed up on. Regular logistic regression, random forests, K-nearest neighbors, classification tree, support vector machine, and gradient boosting were among the algorithms utilized. As a consequence, 2428 responders (71%) completed the second examination after a year. Between baselines, 14% expressed suicidal ideation, resulting in 336 respondents. The gradient boosting technique uses the random forest method as the best method for predicting and predicting suicide attempts. Because of the system's limitations, the amount of data collected was insufficient for complicated algorithms to outperform traditional logistical regression. Including more specific data may improve the accuracy of suicidal behavior and increase machine learning performance over standard approaches. [4]

2.2.3 Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records:

This research serves as a suicide attempt warning system. It was extremely useful for forecasting suicidal behavior and identifying the risk of suicide attempts, as well as analyzing the risk of suicide death. They're working on a system for high-risk suicide attempt patients in this system. They're employed in XGBoost, Local Interpretable Model-agnostic Explanations (LIME), and deep neural networks, among other advanced machine-learning methods. The models were created using electronic health records as a data collection. The purpose was to indicate the likelihood of a suicide attempt over the course of a year. [5]

2.2.4 AI-enabled suicide prediction tools: ethical considerations for medical leaders:

They are attempting to anticipate and reduce suicide attempts using this approach. Artificial intelligence, data science, and other analytical procedures were used. They were attempting to enhance suicide risk assessment and prediction. They are gathering data from two sources: electronic medical records and social media networks. They used two sorts of tools or prediction models: 1. Medical suicide prediction tools attempt to forecast suicide risk

utilizing AI techniques and machine learning with the resources of EMRs, hospital records, and possibly other government data sources. 2. The goal of social suicide prediction tools is to show that AI can use data from social media and browser history to forecast suicide risk. Following the implementation of proper training and education programs, these tools could be put to use. The implementation of these tools may be risk-free as long as they provide public benefit, patient consent, and effective technical security. [6]

2.2.5 Predicting Risk of Suicide Attempts over Time through Machine Learning:

They were attempting to minimize the traditional suicide attempt prediction accuracy and risk detection scale with their approach. The constraints were overcome by applying machine learning to electronic health records using huge medical datasets. These are developed machine learning systems that accurately predict future suicide attempts. They used R, the random Forest package, the ranger program, and logistic regression. The accuracy of this approach improved from 720 days to 7 days before attempting suicide. [7]

2.2.6 Artificial Intelligence and Suicide Prevention: A Systematic Review of Machine Learning:

In order to increase risk detection and examine vast datasets, this system is built on artificial intelligence and machine learning. Using search strings, a review of investigations for suicidal behaviors was conducted using PubMed/MEDLINE, Psych Info, Web-of-Science, and EMBASE. The metrics for analyzing study findings by essential design criteria are area under the curve (AUC), accuracy, sensitivity, and specificity, according to the suicide risk outcome. Incomplete reporting, reporting standards, underreported critical factors, and EMR data are only a few of the drawbacks. The 10 risk factors identified are sleep, circadian, and neural substrates. The use of natural language processing and AI research could have an impact on the prohibition of suicide. [8]

2.2.7 Machine Learning Based Prediction of Suicide Probability:

This technique works by incorporating data classification into various categories into a Bayesian machine learning (ML) model for suicide prediction. Similar techniques such as spline regression are compared to the suggested model. For the dataset used in this study, the model is proven to produce accurate results. The use of Bayesian estimation provides for a larger degree of performance prediction than typical spline regression models, as evidenced by the proposed model's comparatively low root mean square error (RMSE) for all predictions. As a result, the proposed model is highly accurate in calculating the number of suicides, with additional features like the ability to

identify micro trends and the probability risk of individuals according to categories. [9]

2.2.8 Predicting Suicide Attempts and Suicide Deaths Following Outpatient Visits Using Electronic Health Records:

In this system, the author chose to use health records to develop models for predicting suicide attempts and deaths. Prior mental health, suicide attempts, and substance use diagnoses, as well as psychiatric drugs supplied and medical diagnoses, are used to compile the records. According to mental health, the top 5% of people are responsible for 43% of suicide attempts and 48% of suicide deaths. Logistic regression models with penalized LASSO are used for variable selection in a random sample. Both the used two type prediction model and existing suicide prediction technologies outperform one other. [10]

2.2.9 Prediction by data mining, of suicide attempts in Korean adolescents: a national study:

The goal of this project is to implement a suicide attempt prediction model in Korean adolescents. Suicide attempts as the dependent variable is a fixed variable. As independent factors, there are eleven intrapersonal, sociodemographic, and extra personal factors. They attempted to construct a model to predict suicide attempts in Korean youths using decision tree analysis. There are two groups of factors that have been identified. Suicide attempts are 5.4 and 2.8 times greater in depressed people than in non-depressed people. The study's findings suggest that depressed females and those with less familial intimacy be given special attention in order to minimize teenage suicide attempts. [11]

2.2.10 Can acute suicidality be predicted by Instagram data? Results from qualitative and quantitative language analyses:

The relationship between Instagram activity and acute suicidality and language use is investigated in this article. The number of people who contributed to the dataset that was used was 52. Based on photos of self-injury shared on Instagram and a lifelong history of suicidal ideation. On Instagram, text analysis is being used to check for active suicide ideas. Using the Word Count Software and linguistic inquiry, the interviews of text analysis and image captions on Instagram were conducted. To investigate current suicide thoughts, regression analyses are performed using linguistic cues and Instagram activity. According to the conclusions of this study, different machine learning processes may be more useful in comparing used language in interviews. [12]

2.2.11 Predicting future self-harm or suicide in adolescents: a systematic review of risk assessment scales/tools:

The goal is to assess risk instruments' potential to predict future suicide in youth. From conception to 3 March 2018, the data set was extracted from MEDLINE, EMBASE, CUNAHL, and PsycINFO. The data collected on teenagers ranges from 10 to 25 years old. The data was mostly organized by tool and narrative synthesis. A checklist is used to combine the QUIPS (Quality in Prognosis Studies) and QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tools. The setting, demographic, and outcome assessment are all factors in this study. There was an issue with the risk of partiality being unclear, and meta-analysis was not viable due to the large diversity between trials. The purpose of the review is to look into how tools are used in teenagers, however no single tool is suitable for forecasting the risk of suicide or self-harm in teenagers. [13]

2.2.12 Classification of Suicide Attempts through a Machine Learning Algorithm Based on Multiple Systemic Psychiatric Scales:

The goal of the study was to determine the value of avoiding suicide through the classification and prediction of suicide attempts, as well as to see if information gathered from clinical measures may be used to detect real suicide attempts. The contribution of each component was ranked for the classification of suicide attempts using a trained artificial neural network classifier with 41 variables. The model's accuracy results in 93.7 percent detection of suicide attempts within one month, 90.8 percent detection within one year, and 87.4 percent detection across a lifetime. This machine learning technique could assist the rebel identify high-risk patients in clinical definition, depending on the performance of the top five predictors. [14]

2.2.13 Improving the Short-Term Prediction of Suicidal Behavior:

The purpose of this study is to predict who is at danger of trying suicide in the near future. Various questions, such as behavior, have been proposed for the study. To begin, what was the objective of the transition from suicidal thoughts to suicide attempts? Second, how will you make this change in the coming hours, days, or weeks? Third, what is the most significant goal in determining suicidal behavior risk? Fourth, which way of combining data produces the most accurate prediction? These four factors may aid in short-term as well as long-term prediction of suicidal conduct. [15]

2.2.14 Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications:

This article is the first to provide a comprehensive overview of the current suicidal ideation detection (SID) methods. Analyzing data sources, questionnaires, electronic health records, suicide notes, and online user content, the techniques are evaluated. Existing methods such as textual content analysis, such as lexicon-based filtering and word cloud visualization; feature engineering, such as tabular, textual, and affective features; and deep learning-based representation learning, such as CNN- and LSTM-based text encoders, have all been investigated. In the future, SID's main channel 14 will be online social material. As a result, new technologies for detecting online communications carrying suicidal ideation must be developed in the hopes of preventing suicide. [16]

2.2.15 Predicting death by suicide following an emergency department visit for Para suicide with administrative health care system data and machine learning:

The goal of the study is to see if administrative data from the health-care system and machine learning can detect suicide risk in clinical settings. The strategy was developed by comparing prediction models that assess the chance of suicide death with predictors available in health care data. The gradient boosted trees model was used to get a result that might be used in therapeutic settings. The use of administrative data systems in conjunction with personal predictors yielded satisfactory prediction results. [17]

2.2.16 A machine learning approach predicts future risk to suicidal ideation from social media data:

The goal of the system is to create an algorithm called "Suicide Artificial Intelligence Prediction Heuristic (SAIPH)" that can predict future risk of suicidal ideation using data from Twitter. The system was created using a trained set of neural networks that analyzed Twitter data for suicide-related psychological variables. Then, for the output of predict binary suicidal ideation, a random forest model was developed using a neural network. The study's findings suggest that the prediction algorithm has the capacity to forecast individual future suicidal ideation risk and could be used as a clinical decision tool. [18]

2.3 Summary:

We'll talk about existing systems or literature reviews in this chapter. Many researchers are attempting to develop methods that can detect persons who are attempting suicide. We've included some examples of suicide attempt prediction systems as well as their drawbacks. We learn about a variety of existing system models and their flaws.

Chapter-3

Proposed System / Methodology

3.1 Introduction:

The feasibility analysis A Comparative Study on Suicide Attempt Prediction Rate Analysis Based on the Machine Learning Approach system and the requirements for fulfillment of this model are discussed in this chapter. Here is a step-by-step breakdown of our suggested system design. Discuss the architecture's overall details. We demonstrate our proposed system architecture and briefly go over each step. Discuss the proposed system's algorithm and operation procedure as well.

3.2 Feasibility study:

Three researchers and one supervisor were needed for this thesis, which took eight months to complete. Some technical assistance, such as software, is required for the research. The study also needs a dataset and an assessment procedure, both of which are carried out by academics. The data collecting for this thesis is done with the legal feasibility of the dataset in mind. The thesis work did not require any funding from the university or supervisor.

3.3 Requirement Analysis:

In this proposed architecture all requirements here include

a. The computer was very high-performance.
b. We need some open-source software libraries for scientific computations.
c. The open-source software libraries to implement the machine learning model.

3.4 Methodology:

Traditional suicide detection relies on clinical methods, including self-reports and face-to-face interviews. Venek et al. designed a five-item ubiquitous questionnaire for the assessment of suicidal risks and applied a hierarchical classifier on the patients' response to determine their suicidal intentions [19]. In this section, we show the methodology of our proposed model.

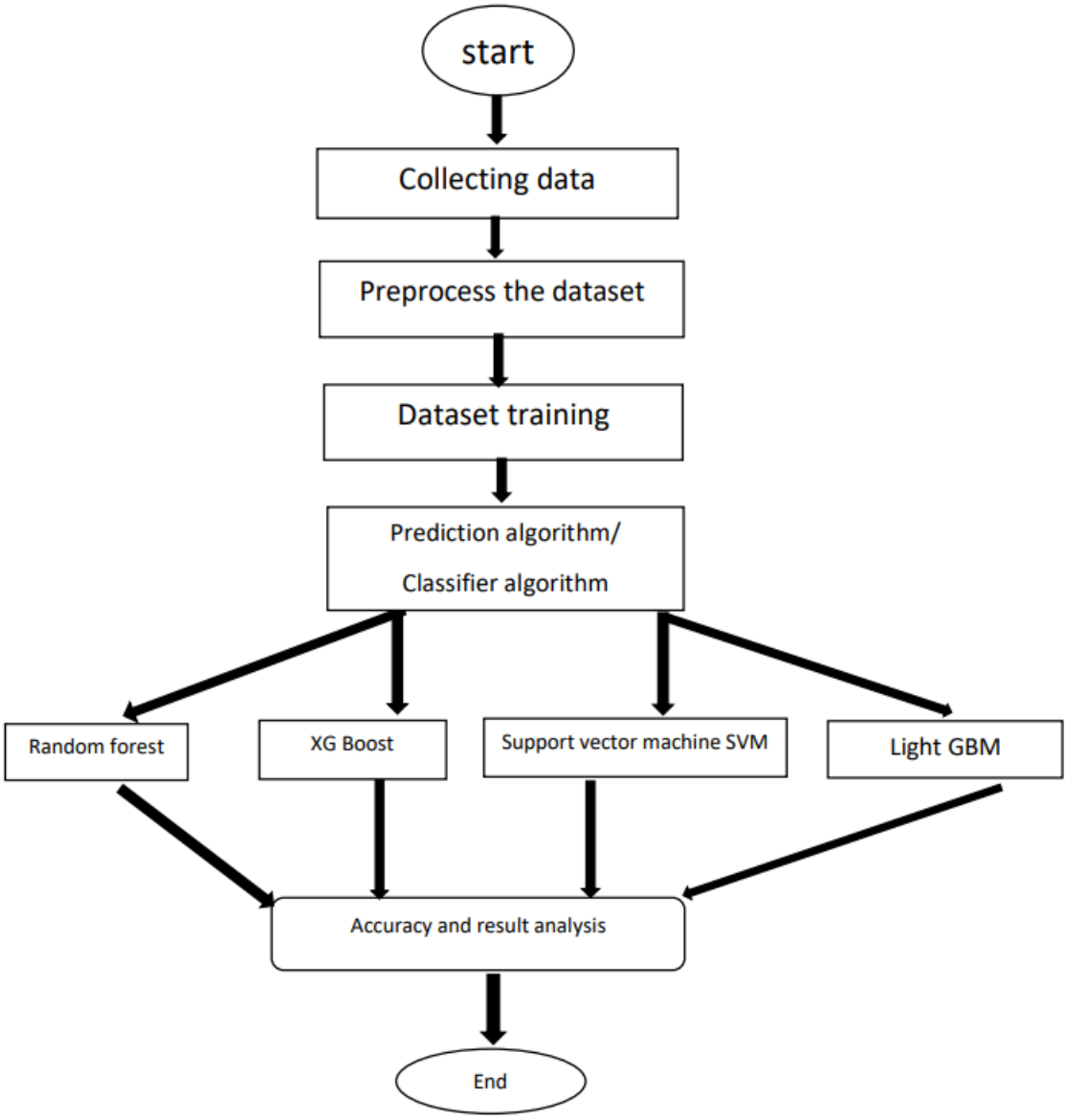


Figure 3.4: Methodology

3.4.1 Collecting data:

Two datasets are used in this system. The first dataset was gathered in Bangladesh, whereas the second was gathered from a variety of different nations and areas. We're gathering the initial data set, which is being gathered by our friends and family. They're all based in Bangladesh. We're gathering information about gender, age, income, body weight, friends, depression, suicide attempts, employment, job title, education level, and how to better oneself. Following the collection of data, we create a CSV file, which is subsequently loaded into our system to train our machine.

3.4.2 Preprocess the dataset:

We are removing several fields from the dataset at this stage. We just take the fields that are required for our experiment. We only evaluated those fields in this experiment that were effective. Our original dataset included 19 variables, but we only utilized our algorithm for 14 of them. For our system, such fields are necessary and effective. We first looked at certain fields that had no impact on our system, then we decided to reduce the number of these fields since they would improve our system's performance while also reducing our time and space complexity. We then reduced those fields and checked our system's result output. We saw that there has no change in our result then finally we fewer those fields this field is no effect on our system output. Then we got better performance output in our system.

3.4.3 Dataset Training:

In this step, we are training a machine to train our dataset. We are using some software and libraries. We are using a python programming language to implement it. At first, we setup our machine. After that, we use Colab for implementation purposes. Here Input some libraries for implementation purposes. Then input our dataset CSV file to train our machine. After that, we use four classifiers whose names are, XG Boost, Light GBM, Support Vector Machine (SVM), the random forest classifier. We were using those classifiers for comparison and prediction purposes. Those classifiers are given some output prediction result, accuracy, recall, precision, classifier error report, and given some other reports. After that, we can analyze this given result.

3.4.4 Prediction algorithm / Classifier algorithm:

We are using four algorithms for comparison and prediction purposes which algorithms name are the Random Forest algorithm, XG Boost algorithm, Support Vector Machine (SVM) algorithm, Light GBM algorithm. Discuss our algorithms detailed below.

3.4.4.1 Random Forest Algorithm:

The Random Forest algorithm is a supervised learning algorithm. Random forest is used in classification and regression analysis. Random Forest generates multiple decision trees for a data sample, then merges the results of those trees, predicts the outcome, and selects the best result. Random Forest is an ensemble method. The random forest algorithm's working procedure, the forest it can construct, is an ensemble of decision trees, and it can typically have trained the bagging method. The bagging method is a combination of learning models that increases the overall result. One big advantage of this algorithm can be used for both classification and regression problems. It can overcome the overfitting problem by averaging or combining the results of different decision trees. It also works for a broader set of data items than a single decision tree. Has less variance than a single decision tree. It is very flexible and possesses very high accuracy. It maintains good accuracy even after providing data without scaling. It can maintain good accuracy even if a large proportion of the data is missing. It has some disadvantages; complexity is the main disadvantage. Its construction is much harder and time-consuming than decision trees. More computational resources are required. It is less intuitive in the case when we have a large collection of decision trees. Its prediction process is very time-consuming in comparison with other algorithms.

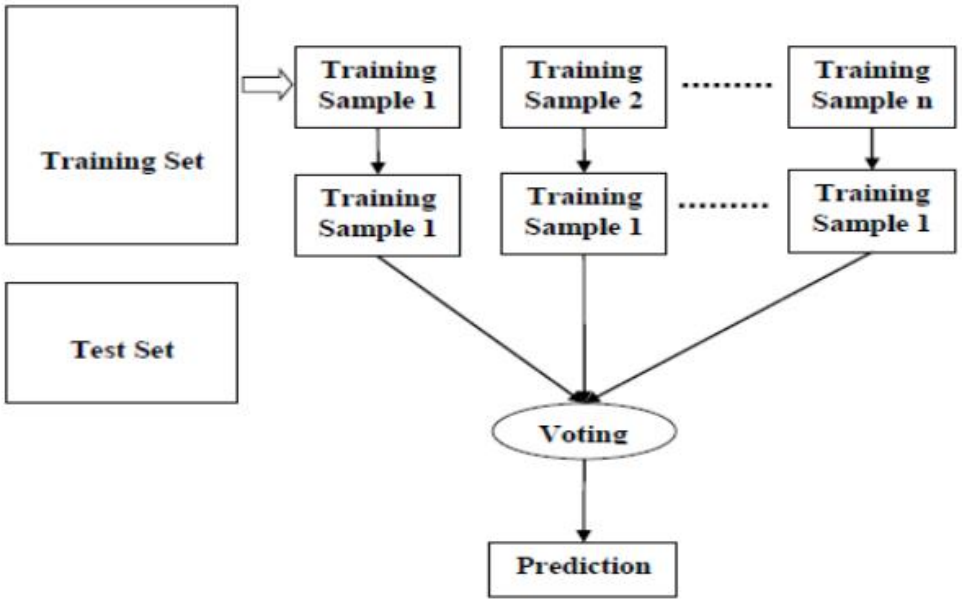


Figure 3.4.4.1: Random Forest algorithm

3.4.4.2 XG Boost algorithm:

XG Boost is very popular and most used algorithm in machine learning. XG Boost is a decision tree base ensemble algorithm. It is used in gradient boosted trees which is designed by speed and performance. The library is laser focused on computational speed and model performance. XG Boost model features the implementation of the model supports the features of the scikit-learn and R implementations with new additions regularization. Three main forms of gradient boosting are supported: Gradient Boosting, Stochastic Gradient Boosting, Regularized Gradient Boosting. Its system features have parallelization, Distributed computing, out-of-core computing, cache optimization. Its algorithm features have Sparse Aware, Block Structure, Continued Training. Mainly two reasons to use XG Boost, 1. Execution speed, 2. Model performance. It is very fast algorithm. It is dominating structured or tabular datasets on classification and regression predictive modeling problems.

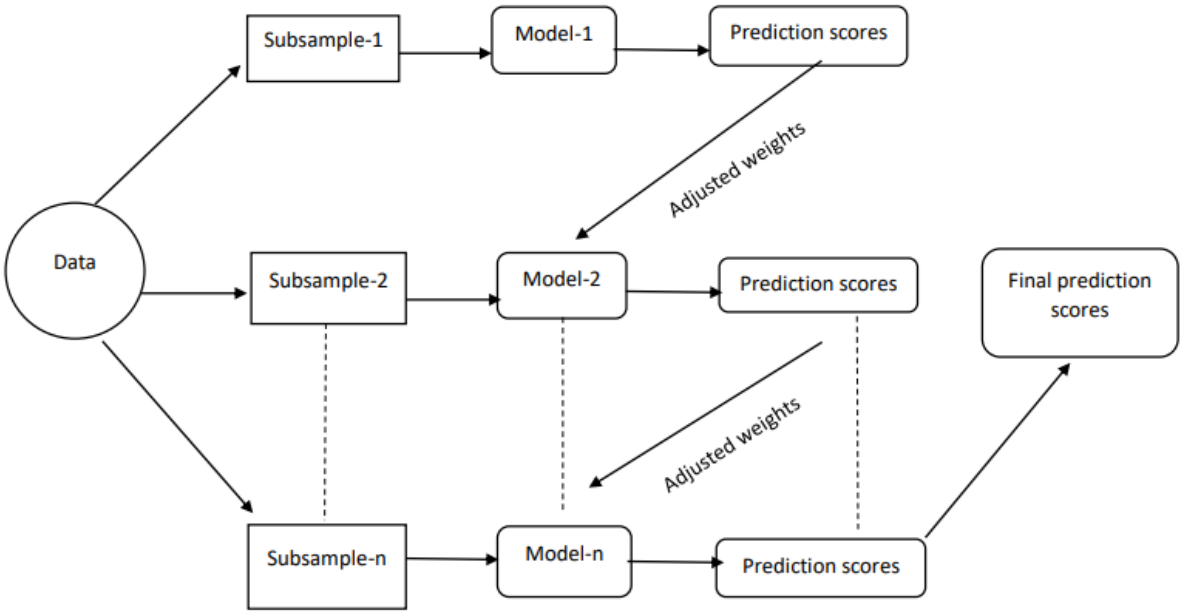


Figure 3.4.4.2: XG Boost algorithm

3.4.4.3 Support vector machine SVM:

Support Vector Machine SVM is most widely used algorithm. It's a supervised machine learning algorithm. SVM is used for classification and regression problems but generally, it is used in classification problems. An SVM model is working, we plot data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. After that, we perform classification by finding the hyper-plane that differentiates the two classes very well. The hyperplane will be generated iteratively by SVM so that the error can be minimized. Support vectors are simply the coordinates of individual observation. SVM classifier offers great accuracy and works well with high dimensional space. Its classifier uses a subset of training points hence in result uses very little memory. It has high training time hence in practice not suitable for large Data Subsample-1 Subsample-2 Subsample-n Model-1 Model-2 Model-n Prediction scores Prediction scores Prediction scores Final prediction scores-23 datasets. It is that SVM classifiers do not work well with overlapping classes. SVM are powerful tools, but their compute and storage requirements increase rapidly with the number of training vectors.

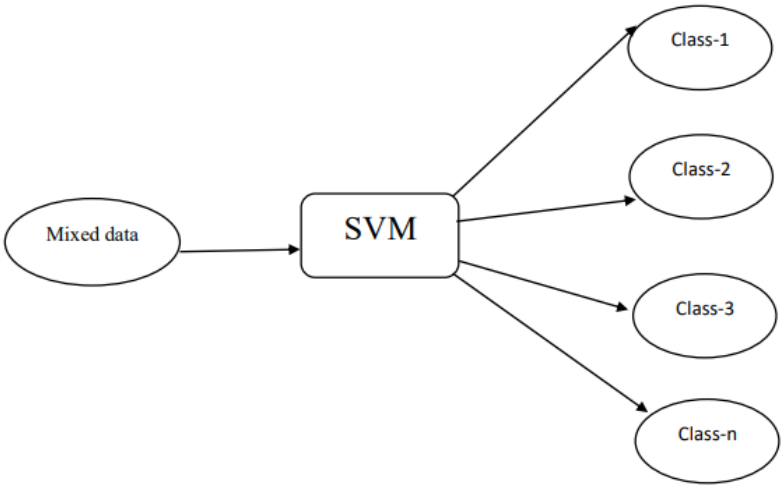


Figure 3.4.4.3.1: Support vector machine SVM

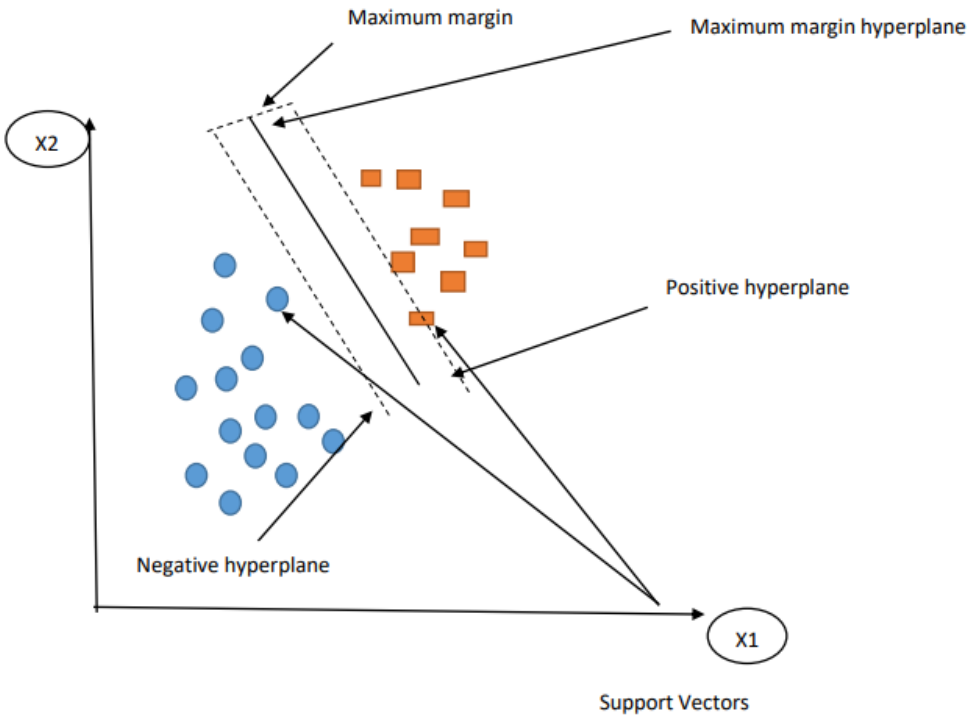


Figure 3.4.4.3.2: Support vector machine SVM

3.4.4.4 Light GBM algorithm:

Light GBM is a fast, distributed, high-performance gradient boosting framework based a decision tree algorithm. It is used for ranking, classification, and many other machine learning tasks. It is based on a decision tree algorithm. In this tree splits leaf wise. The leaf wise algorithm can reduce more loss than the level-wise algorithm and much better accuracy the result in other boosting algorithms. LightGBM here Light because of its computation power and giving results faster. It takes less memory to run and can deal with large amounts of data. It is a very fast processing algorithm. It has more than 100 parameters. Light GBM has some advantages which are faster training speed and higher efficiency, Lower memory usage, better accuracy than any other boosting algorithm, Compatibility with large datasets, Parallel learning supported.

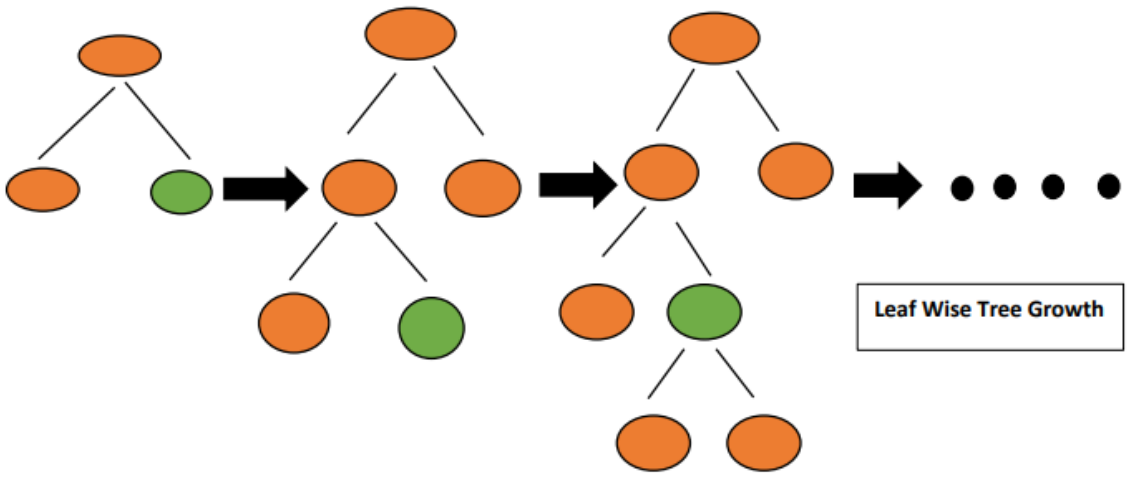


Figure 3.4.4.4: Light GBM algorithm

3.5 Summary:

We discuss a proposed model architecture in this chapter that we employed in the "Suicide Attempt Prediction Rate Analysis Using Machine Learning Approach." We're just going through the basics of the system. We'll talk about our classifiers, which include the Random Forest algorithm, XG Boost, Support Vector Machine SVM, and Light GBM algorithm.

Chapter-4

Experimental Result Analysis

4.1 Introduction:

The first stage in the analysis of quantitative data is to organize the raw data in a way that makes them more easily understood [20]. In this chapter, we will go over our experimental results analysis. We're displaying a data visualization of our whole system. We present existing data experimental result analysis, our own data experimental result analysis, and lastly merged data experimental result analysis here.

4.2 Experimental Result Analysis for Existing dataset:

4.2.1 Plotting gender data:

We're plotting gender statistics here. We demonstrate how often men, women, and transgender people are in the dataset.

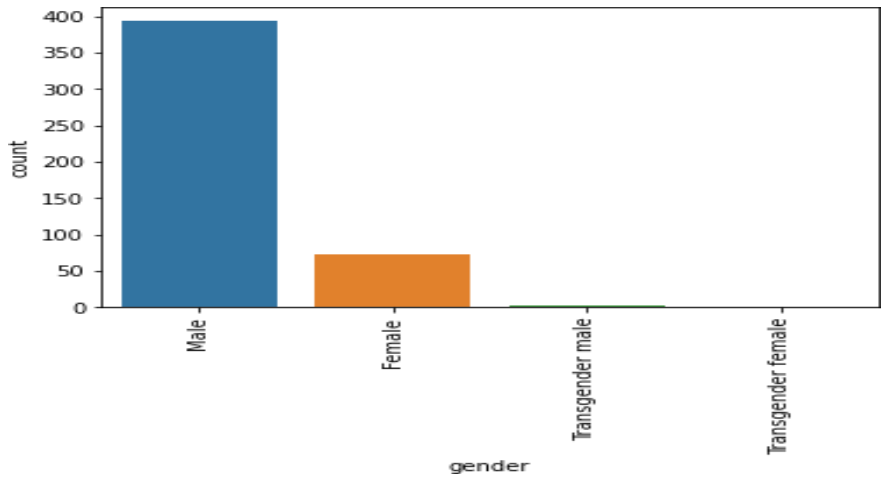


Figure 4.2.1: Plotting Gender data

4.2.2 Plotting age data:

We're plotting age data here. We show data on suicide attempters' ages. We can see that people of various ages are attending this survey.

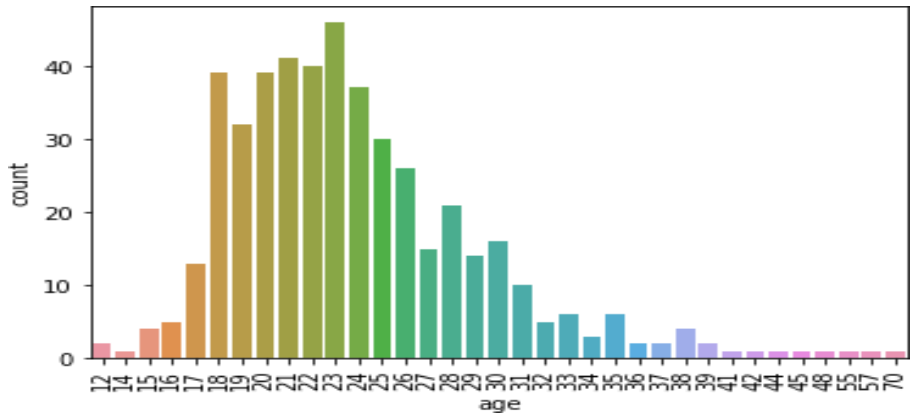


Figure 4.2.2: Plotting age data

4.2.3 Plotting Income Data:

Here we are plotting income data. We present income data for participants of this survey. We can observe various forms of income statistics.

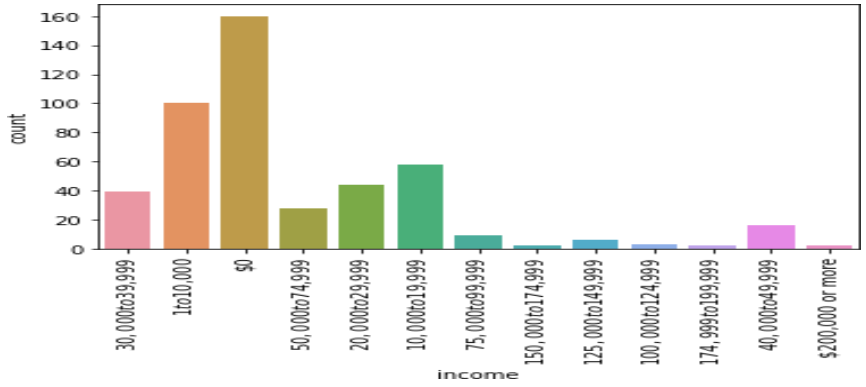


Figure 4.2.3: Plotting Income Data

4.2.4 Plotting Bodyweight data:

Bodyweight data is plotted here. We display the Bodyweight data of those who have included in dataset.

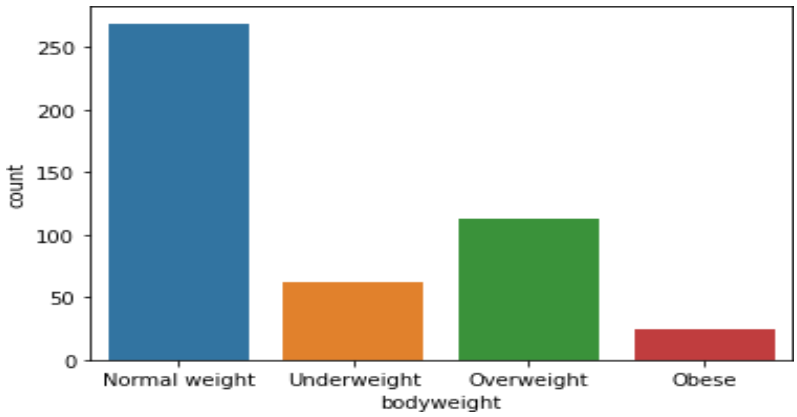


Figure 4.2.4: Plotting Bodyweight data

4.2.5 Plotting Virginity data:

Here we are plotting Virginity data. We show that Dataset participants Virginity data. We see how many people are virgins or not virgins in this dataset.

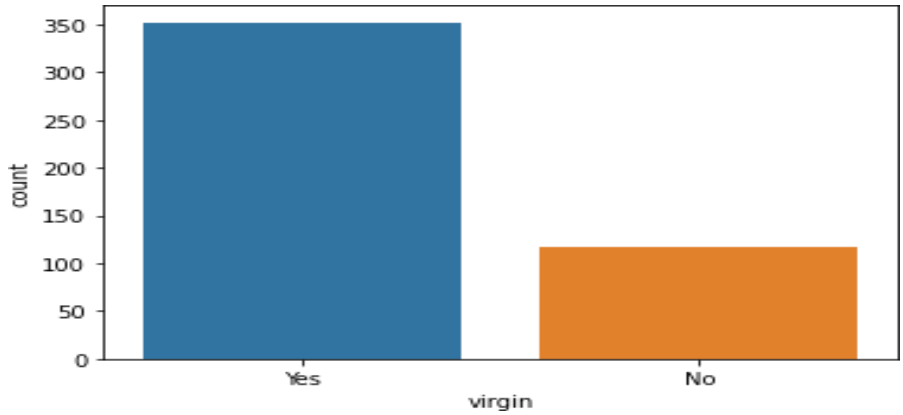


Figure 4.2.5: Plotting Virginity data

4.2.6 Plotting Social Fear data:

Here we are plotting Social Fear data. we show that Dataset participants people Social Fear data state.

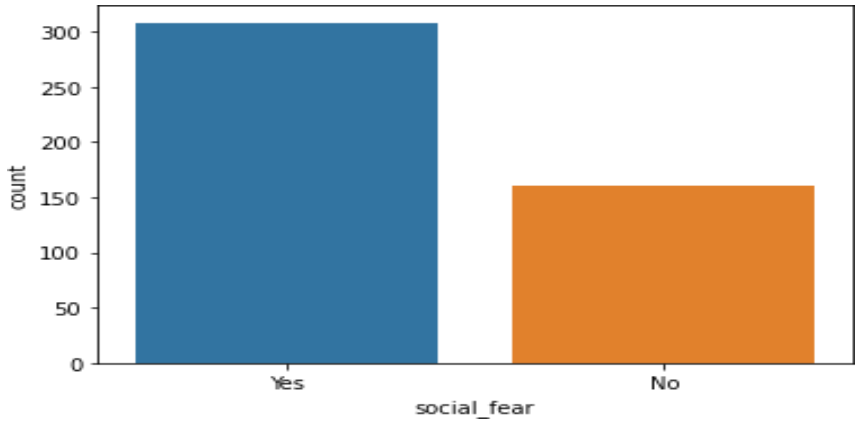


Figure 4.2.6: Plotting Social Fear data

4.2.7 Plotting depressed data:

Here we are plotting Depression state data. We show here Dataset participants depression state. Here wesee how many people are depressed or not.

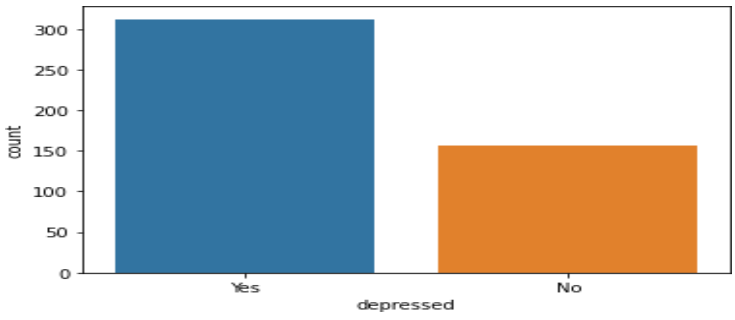


Figure 4.2.7: Plotting Depression state data

4.2.8 Plotting Attempt Suicide data:

Here we are plotting Dataset participants' Suicide Attempt data. We can see how many people are Attempt Suicide or not.

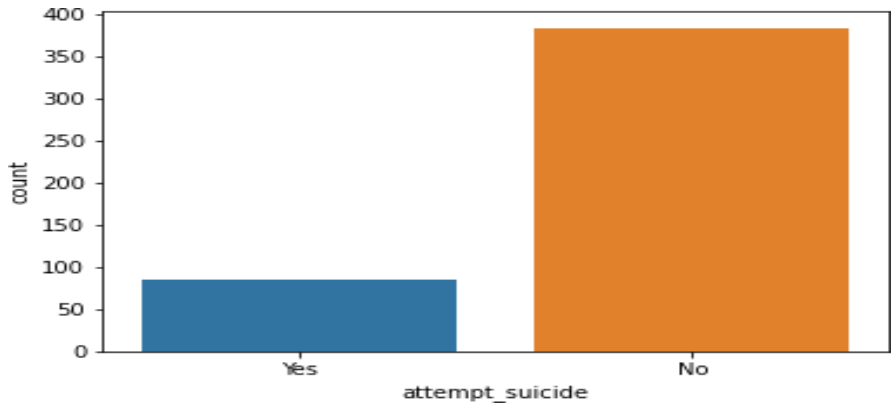


Figure 4.2.8: Plotting Attempt Suicide data

4.2.9 Plotting what help from others data:

Here we are plotting what types of help from others data. We show that suicide attempt people what help from others. We see how many people are helped from others.

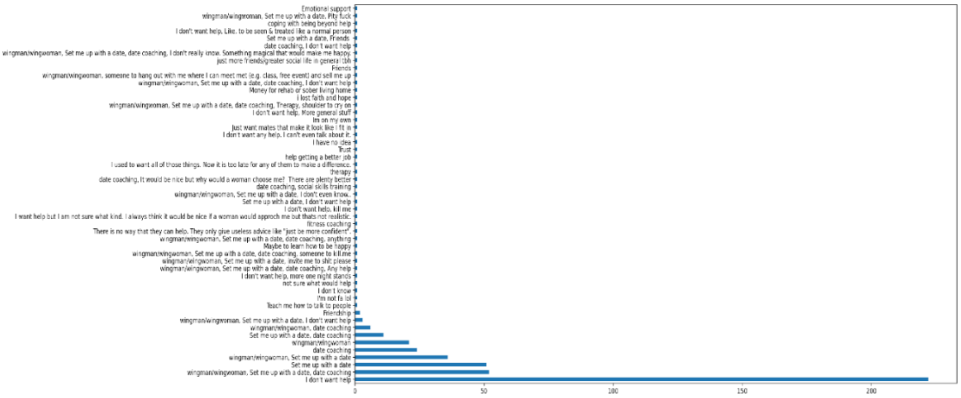


Figure 4.2.9: Plotting What help from other data

4.2.10 XG Boost Classifier:

4.2.10.1 Plotting XG Boost log loss:

Here we are plotting XG Boost log loss. We are showing how many data are lost with respect to train and test data.

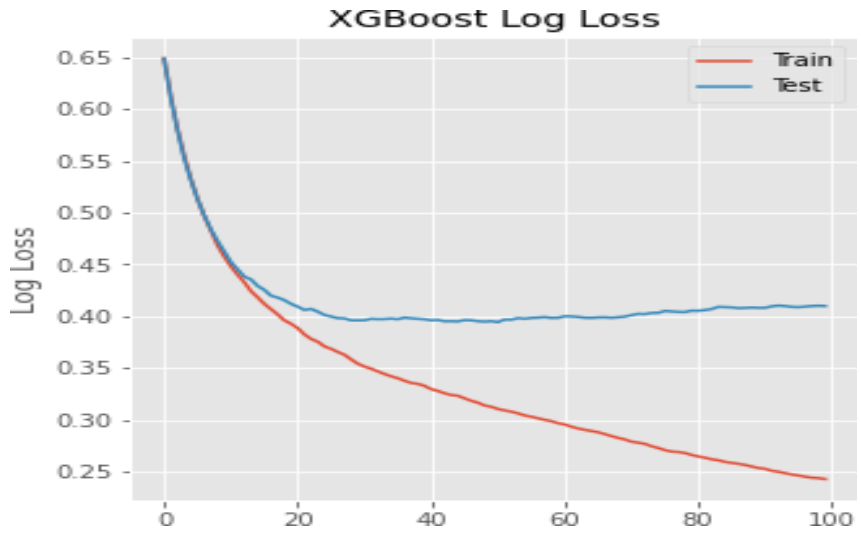


Figure 4.2.10.1: Plotting XG Boost log loss

4.2.10.2 Plotting XG Boost Classification error:

Here we are plotting XG Boost Classification error. We are showing how many data are Error with respect to train and test data.

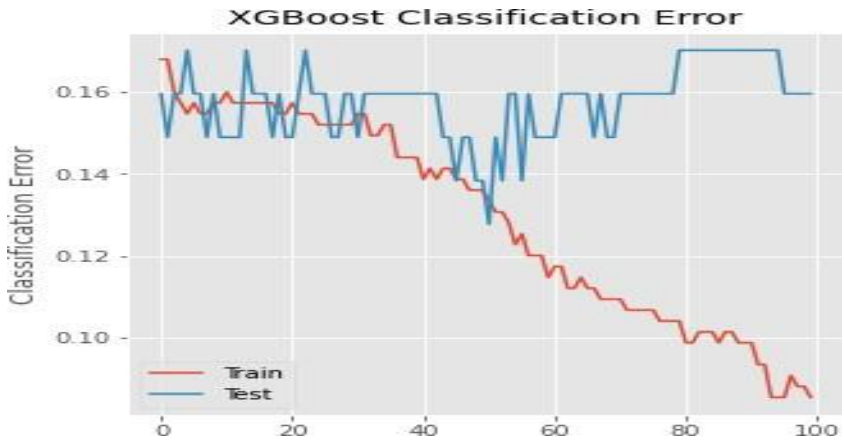
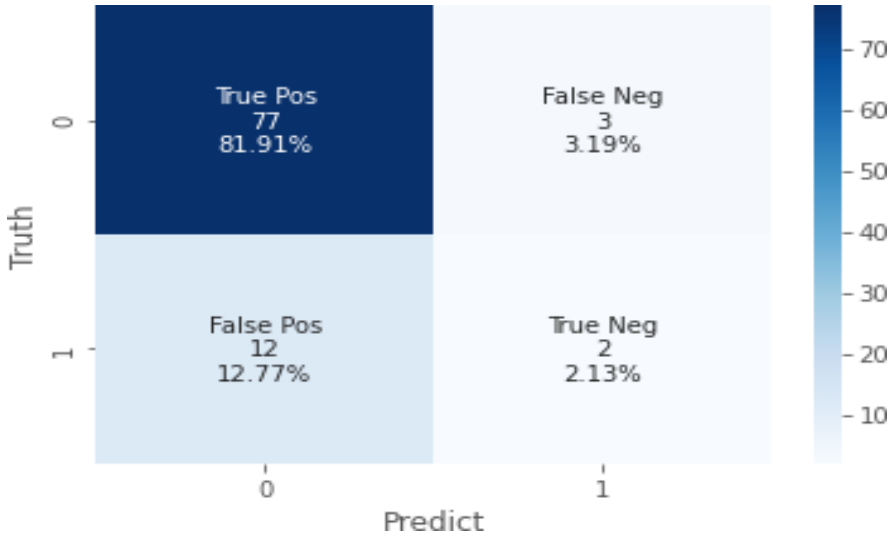


Figure 4.2.10.2: Plotting XG Boost Classification error

4.2.10.3 Plotting XG Boost Classification prediction report:

Here we are plotting XG Boost Classification report. In this plot, we are showing how many data are actually true and false detected with respect to train and test data and also show their prediction result.



	precision	recall	f1-score	support
0	0.87	0.96	0.91	80
1	0.40	0.14	0.21	14
accuracy			0.84	94
macro avg	0.63	0.55	0.56	94
weighted avg	0.80	0.84	0.81	94

Figure 4.2.10.3: Plotting XG Boost Classification prediction report

4.2.10.4 Plotting XG Boost Classifier True and False Positive Rate:

Here we are plotting XG Boost Classifier True and False Positive Rate. In this plot we are showing how many data are actually True and false detection rate with respect to train and test data.

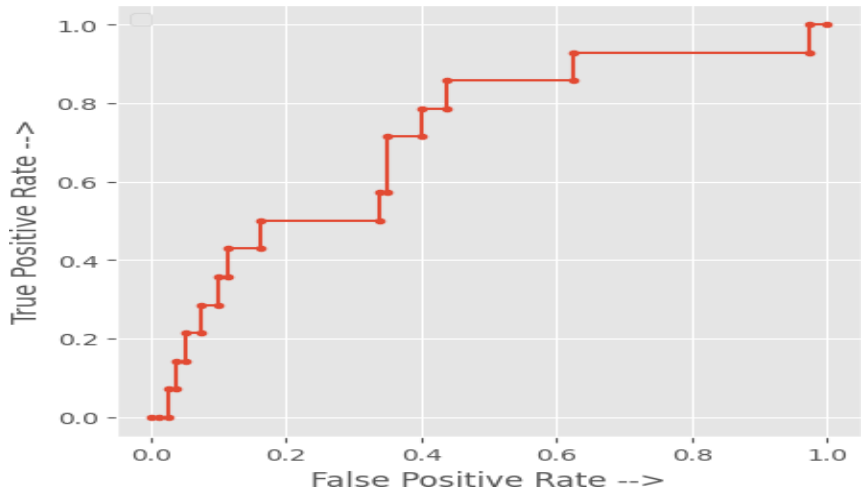


Figure 4.2.10.4: Plotting XG Boost Classifier True and False Positive Rate

4.2.11 Plotting LightGBM Classifier:

4.2.11.1 Plotting LightGBM Classifier Metric During Training:

Here we are showing LightGBM classifier Metric During Training with respect to test and train dataset.

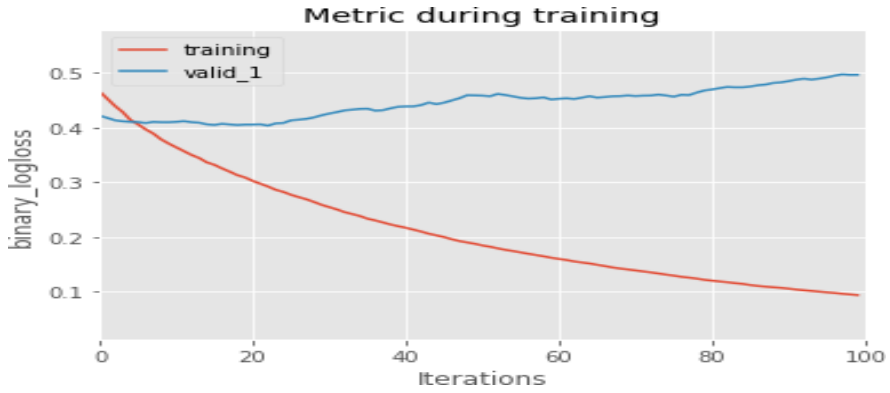
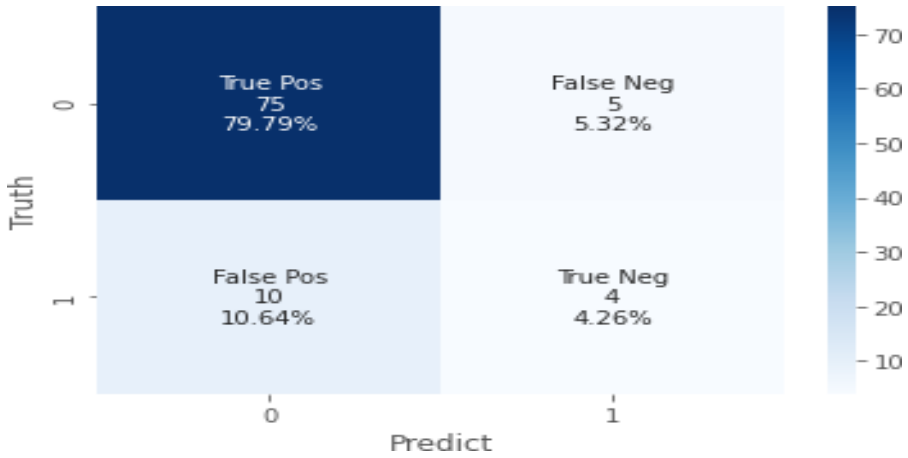


Figure 4.2.11.1: Plotting LightGBM Classifier Metric During Training

4.2.11.2 Plotting LightGBM Classification report:

Here we are plotting LightGBM Classification Report. In this plot we are showing how many data are actually true and false detected with respect to train and test data and also show their prediction result.



	precision	recall	f1-score	support
0	0.88	0.94	0.91	80
1	0.44	0.29	0.35	14
accuracy			0.84	94
macro avg	0.66	0.61	0.63	94
weighted avg	0.82	0.84	0.83	94

Figure 4.2.11.2: Plotting LightGBM Classification report

4.2.11.3 Plotting LightGBM Classifier True and False Positive Rate:

Here we are plotting LightGBM Classifier True and False Positive Rate. In this plot we are showing how many data are actually True and false detection rate with respect to train and test data.

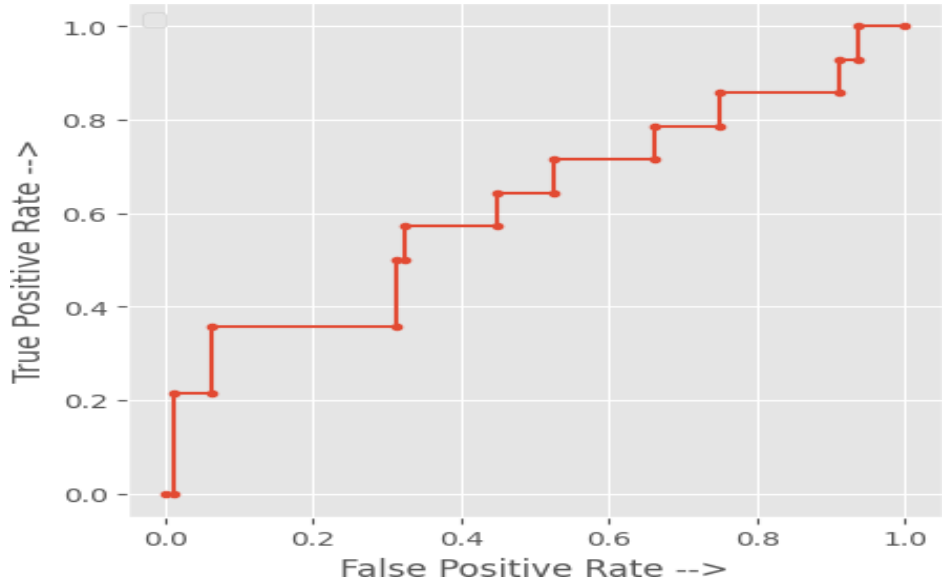
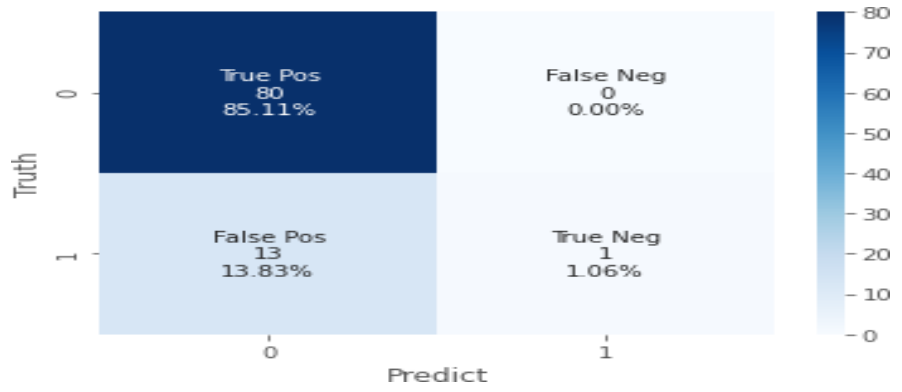


Figure 4.2.11.3: Plotting LightGBM Classifier True and False Positive Rate

4.2.12 Plotting Support Vector Machine (SVM) Classifier:

4.2.12.1 Plotting Support Vector Machine (SVM) Classification report:

Here we are plotting Support Vector Machine (SVM**)** Classification Report. In this plot we are showing how many data are actually true and false detected with respect to train and test data and also show their prediction result.



	precision	recall	f1-score	support
0	0.86	1.00	0.92	80
1	1.00	0.07	0.13	14
accuracy			0.86	94
macro avg	0.93	0.54	0.53	94
weighted avg	0.88	0.86	0.81	94

Figure 4.2.12.1: Plotting Support Vector Machine (SVM) Classification report

4.2.12.2 Plotting Support Vector Machine (SVM) Classifier True and False Positive Rate:

Here we are plotting Support Vector Machine (SVM**)** Classifier True and False Positive Rate. In this plot we are showing how many data are actually True and false detection rate with respect to train and test data.

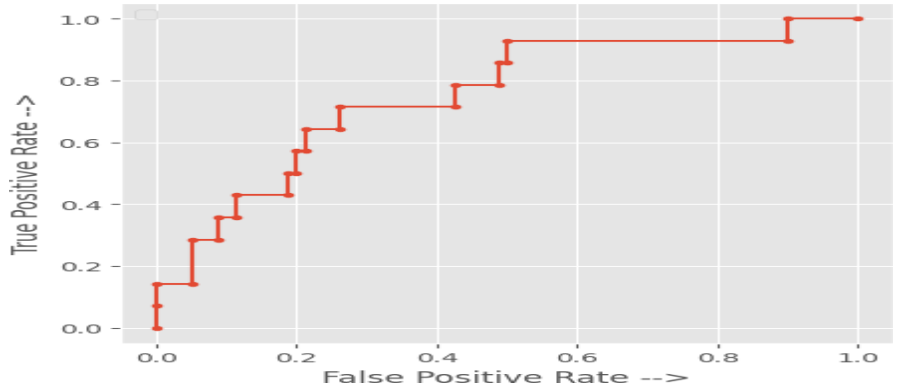
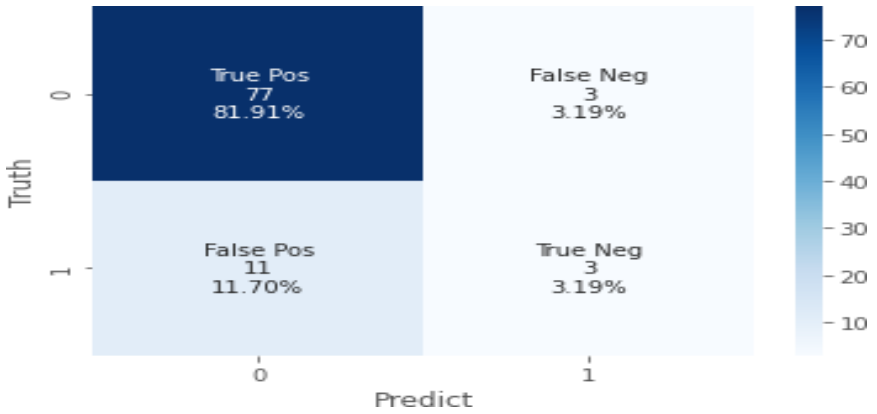


Figure 4.2.12.2: Plotting Support Vector Machine (SVM) Classifier True and False Positive Rate

4.2.13 Plotting Random Forest Classifier:

4.2.13.1 Plotting Random Forest Classification report:

Here we are plotting Random Forest Classification Report. In this plot we are showing how many data are actually true and false detected with respect to train and test data and also show their prediction result.



	precision	recall	f1-score	support
0	0.88	0.96	0.92	80
1	0.50	0.21	0.30	14
accuracy			0.85	94
macro avg	0.69	0.59	0.61	94
weighted avg	0.82	0.85	0.82	94

Figure 4.2.13.1: Plotting Random Forest Classification report

4.2.13.2 Plotting Random Forest Classifier True and False Positive Rate:

Here we are plotting Random Forest Classifier True and False Positive Rate. In this plot we are showing how many data are actually True and false detection rate with respect to train and test data.

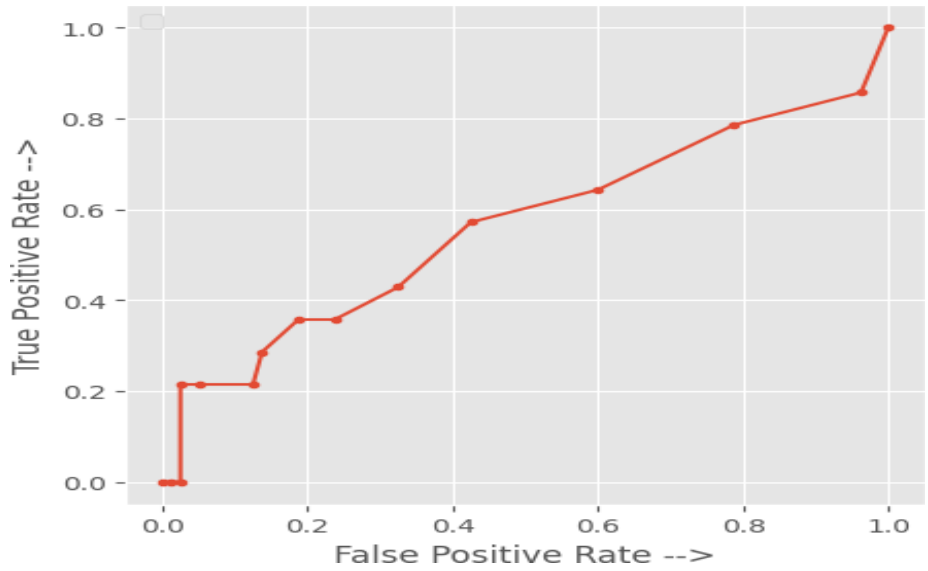


Figure 4.2.13.2: Plotting Random Forest Classifier True and False Positive Rate

4.2.13.3 ROC curve (receiver operating characteristic curve):

Here we are applying ROC curve on our Brazilian dataset according to our four algorithms.

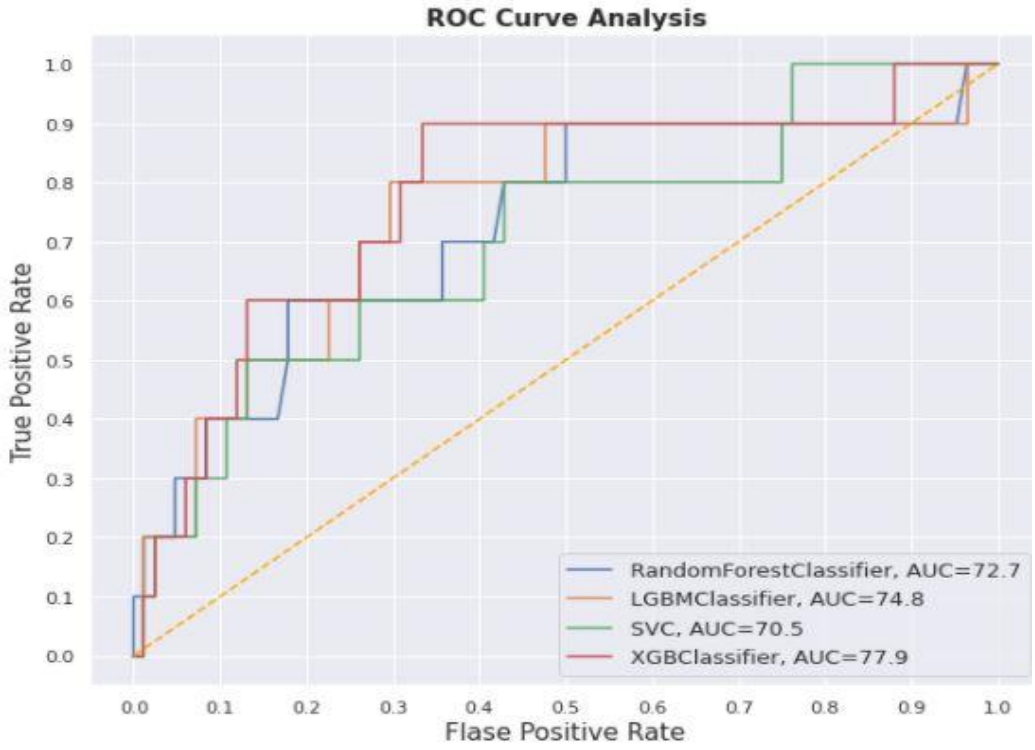


Figure 4.2.13.3: ROC Curve Analysis

4.3 Experimental Result Analysis for Our Own dataset:

4.3.1 Plotting Gender data:

Here we are plotting gender data. We show that how many male & female people they are attending for collecting dataset.

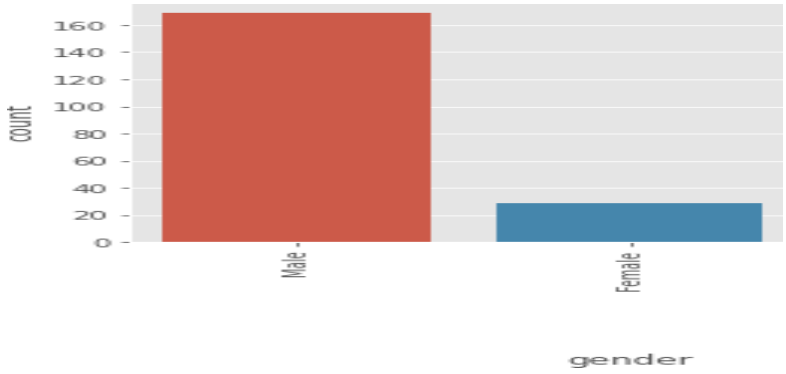


Figure 4.3.1: Plotting Gender data

4.3.2 Plotting age data:

Here we are plotting age data. We show that suicide attempt people age data. Here we can see that different types of age people they attended for collecting our dataset.

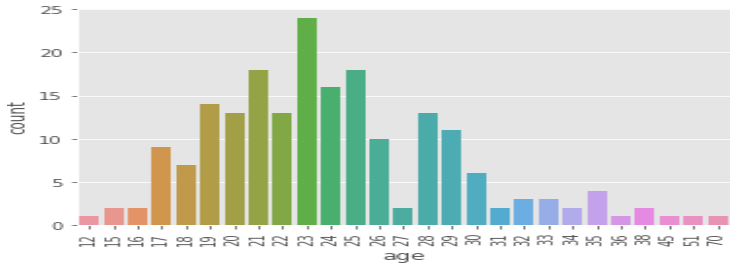


Figure 4.3.2: Plotting age data

4.3.3 Plotting Income Data:

Here we are plotting income data. We show the income data of the people who attended for collecting our dataset. We see that different types of income data here.

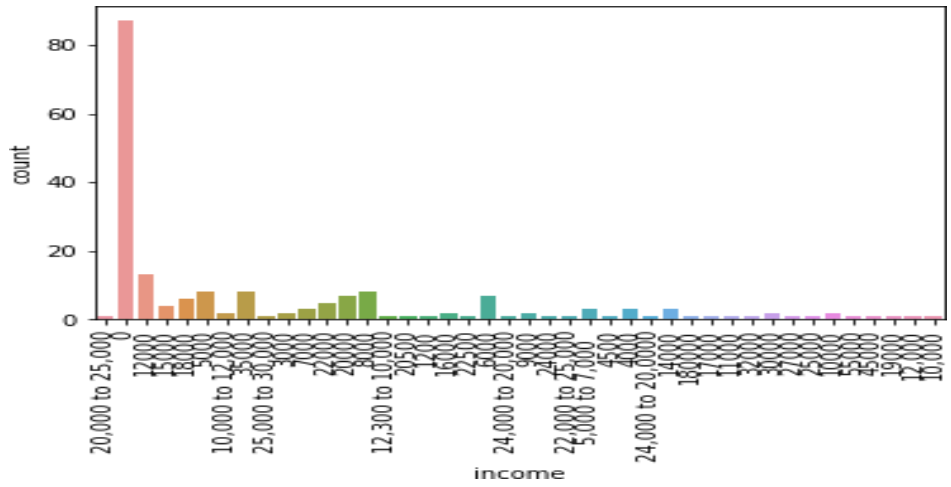


Figure 4.3.3: Plotting Income Data

4.3.4 Plotting Bodyweight data:

Here we are plotting Bodyweight data. We can see the bodyweight data of the people who were attended for collecting dataset.

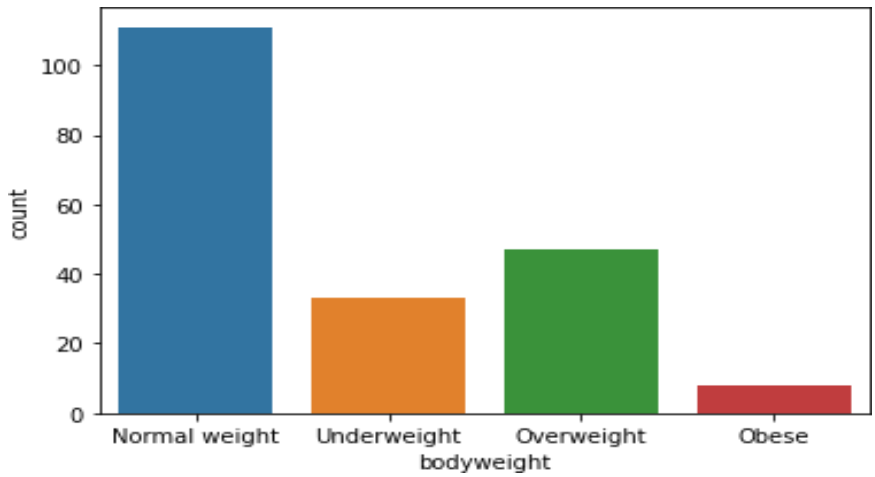


Figure 4.3.4: Plotting Bodyweight data

4.3.5 Plotting Virginity data:

Here we are plotting Virginity data. We can see the Virginity data of the people who attended for collecting our dataset. We see that how many people are virgin or not virgin.

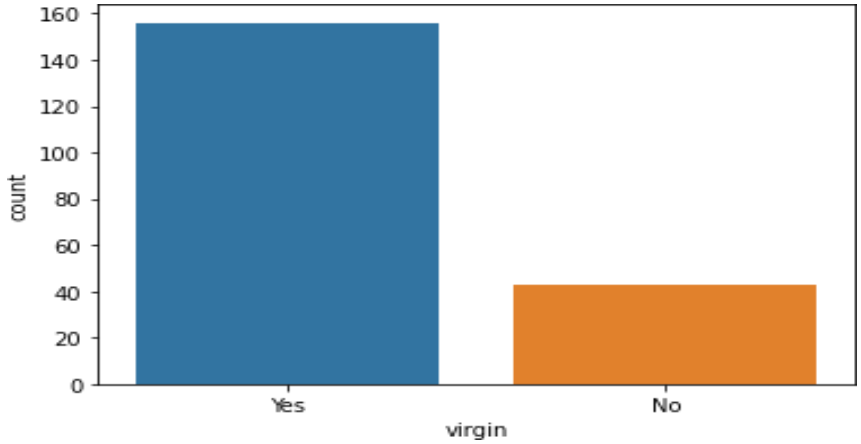


Figure 4.3.5: Plotting Virginity data

4.3.6 Plotting Social Fear data:

Here we are plotting Social Fear data. We show the Social Fear data of the people whom attended for collecting our dataset.

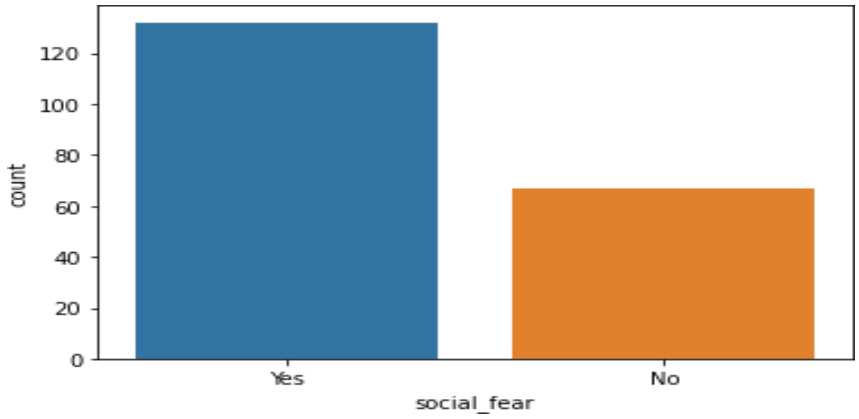


Figure 4.3.6: Plotting Social Fear data

4.3.7 Plotting depressed data:

Here we are plotting depression state data. We can see that depression state of the people whom attended for our dataset. That suicide attempt people. We see that how many people are depressed or not depressed.

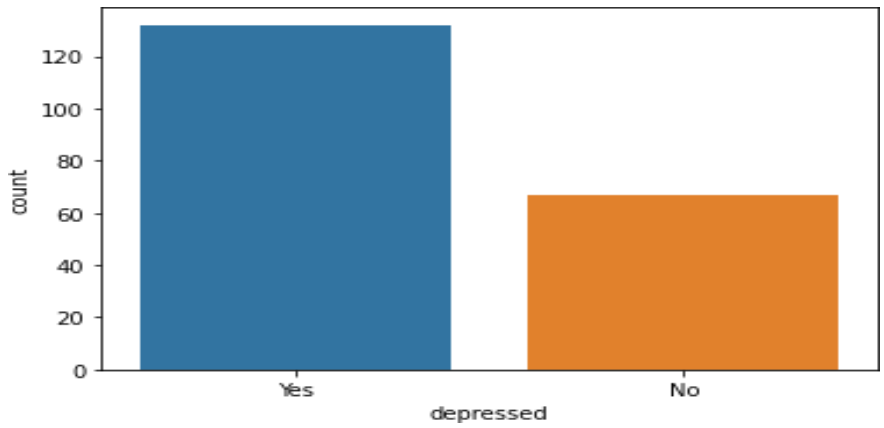


Figure 4.3.7: Plotting depressed data

4.3.8 Plotting Attempt Suicide data:

Here we are plotting Attempt Suicide data. We see that how many people are Attempt Suicide or not Attempt Suicide.

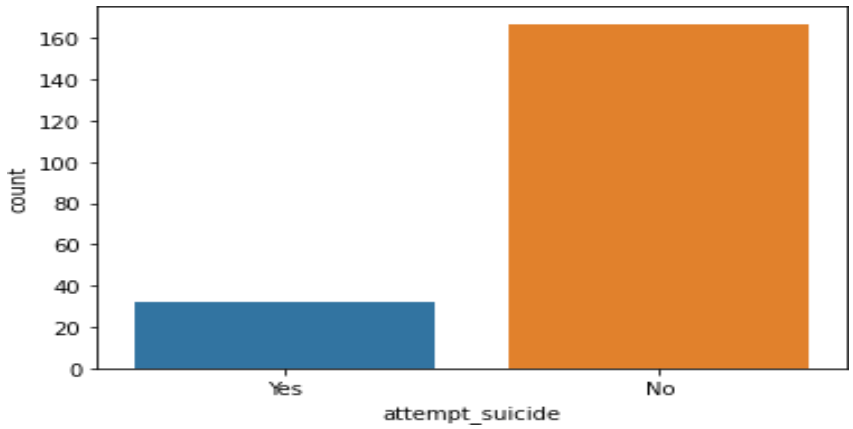


Figure 4.3.8: Plotting Attempt Suicide data

4.3.9 Plotting what help from others data:

Here we are plotting what help from others data. We show that suicide attempt people what help from others. We see that how many people are help from others.

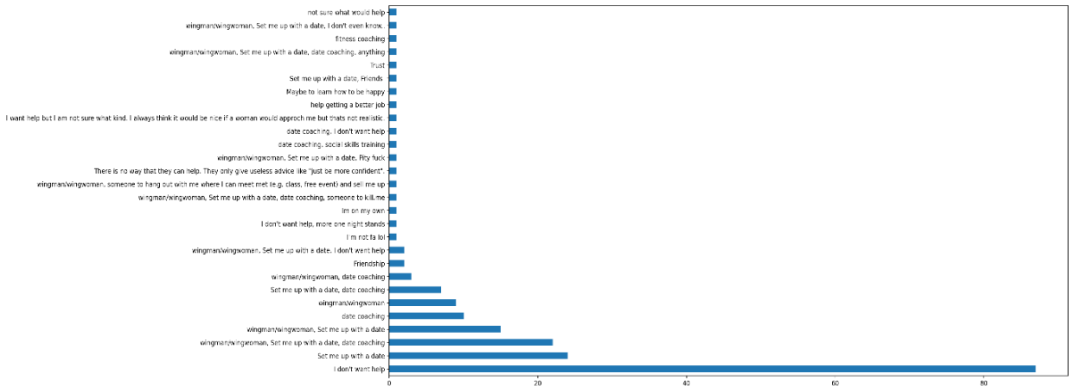


Figure 4.3.9: Plotting what help from others data

4.3.10 XG Boost Classifier:

4.3.10.1 Plotting XG Boost log loss:

Here we are plotting XG Boost log loss. We are showing how many data are losses with respect to train and test data.

Figure 4.3.10.1: Plotting XG Boost log loss

4.3.10.2 Plotting XG Boost Classification error:

Here we are plotting XG Boost Classification error. We are showing how many data are Error with respect to train and test data.

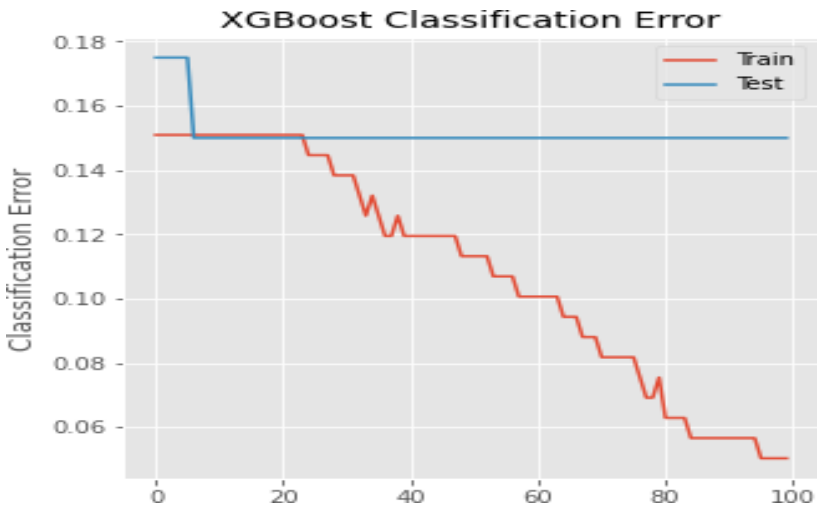
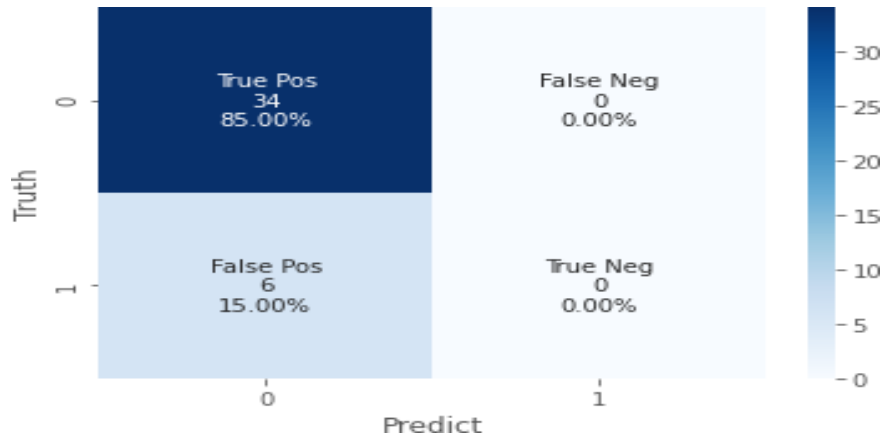


Figure 4.3.10.2: Plotting XG Boost Classification error

4.3.10.3 Plotting XG Boost Classification prediction report:

Here we are plotting XG Boost Classification report. In this plot we are showing how many data are actually true and false detected with respect to train and test data and also show their prediction result.



	precision	recall	f1-score	support
0	0.85	1.00	0.92	34
1	0.00	0.00	0.00	6
accuracy			0.85	40
macro avg	0.42	0.50	0.46	40
weighted avg	0.72	0.85	0.78	40

Figure 4.3.10.3: Plotting XG Boost Classification prediction report

4.3.10.4 Plotting XG Boost Classifier True and False Positive Rate:

Here we are plotting XG Boost Classifier True and False Positive Rate. In this plot we are showing how many data are actually true and false detection rate with respect to train and test data.

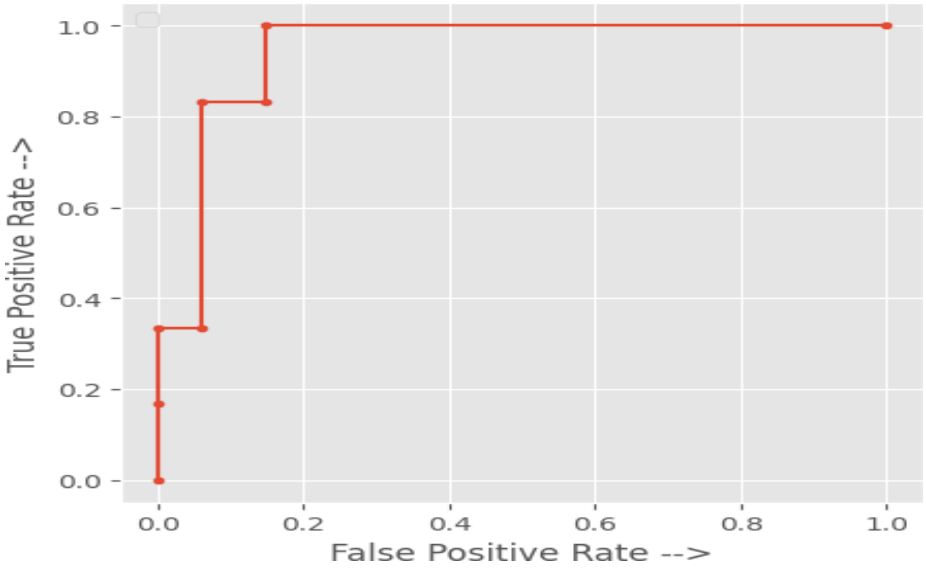


Figure 4.3.10.4: Plotting XG Boost Classifier True and False Positive Rate

4.3.11 Plotting LightGBM Classifier:

4.3.11.1 Plotting LightGBM Classifier Metric During Training:

Here we are showing LightGBM classifier Metric during Training with respect to test and train dataset.

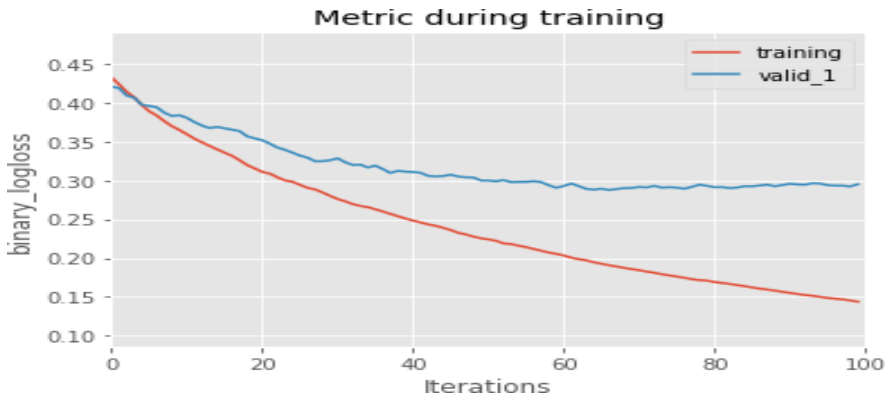
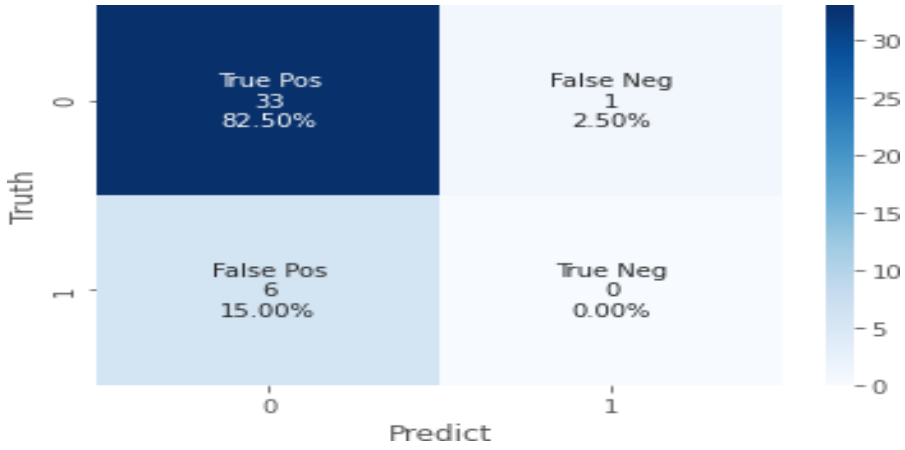


Figure 4.3.11.1: Plotting LightGBM Classifier Metric during Training

4.3.11.2 Plotting LightGBM Classification report:

Here we are plotting LightGBM Classification Report. In this plot we are showing how many data are actually true and false detected with respect to train and test data and also show their prediction result.



	precision	recall	f1-score	support
0	0.85	0.97	0.90	34
1	0.00	0.00	0.00	6
accuracy			0.82	40
macro avg	0.42	0.49	0.45	40
weighted avg	0.72	0.82	0.77	40

Figure 4.3.11.2: Plotting LightGBM Classification report

4.3.11.3 Plotting LightGBM Classifier True and False Positive Rate:

Here we are plotting LightGBM Classifier True and False Positive Rate. In this plot we are showing how many data are actually true and false detection rate with respect to train and test data.

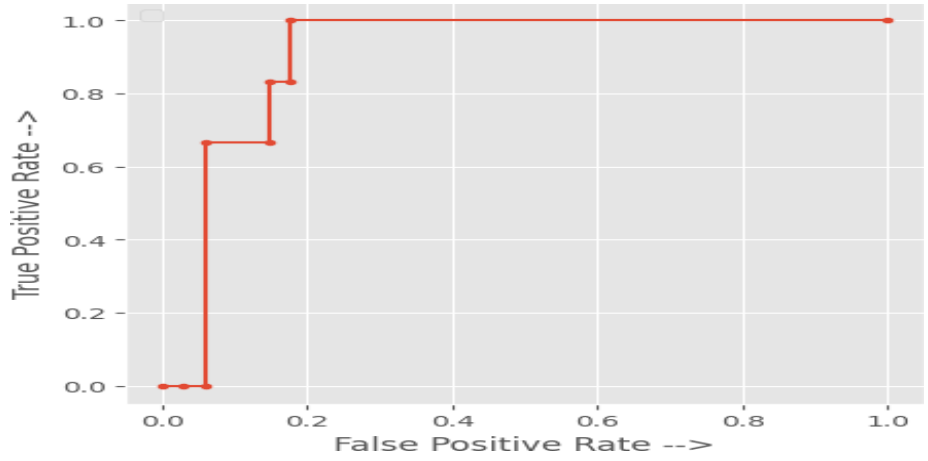
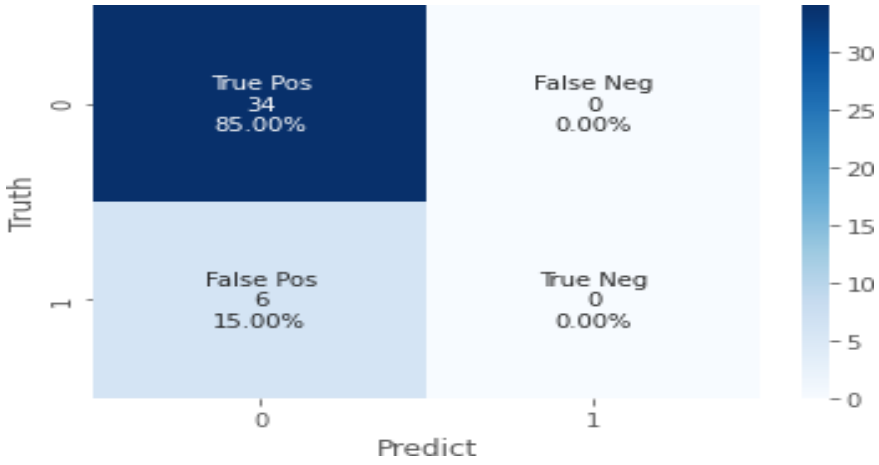


Figure 4.3.11.3: Plotting LightGBM Classifier True and False Positive Rate

4.3.12 Plotting Support Vector Machine (SVM) Classifier:

4.3.12.1 Plotting Support Vector Machine (SVM) Classification report:

Here we are plotting Support Vector Machine (SVM**)** Classification Report. In this plot we are showing how many data are actually true and false detected with respect to train and test data and also show their prediction result.



	precision	recall	f1-score	support
0	0.85	1.00	0.92	34
1	0.00	0.00	0.00	6
accuracy			0.85	40
macro avg	0.42	0.50	0.46	40
weighted avg	0.72	0.85	0.78	40

Figure 4.3.12.1: Plotting Support Vector Machine (SVM) Classification report

4.3.12.2 Plotting Support Vector Machine (SVM) Classifier True and False Positive Rate:

Here we are plotting Support Vector Machine (SVM**)** Classifier True and False Positive Rate. In this plot we are showing how many data are actually True and false detection rate with respect to train and test data.

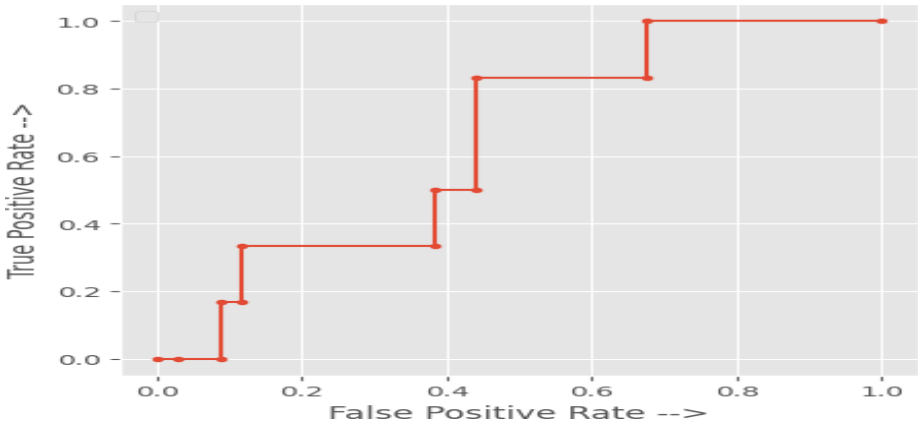
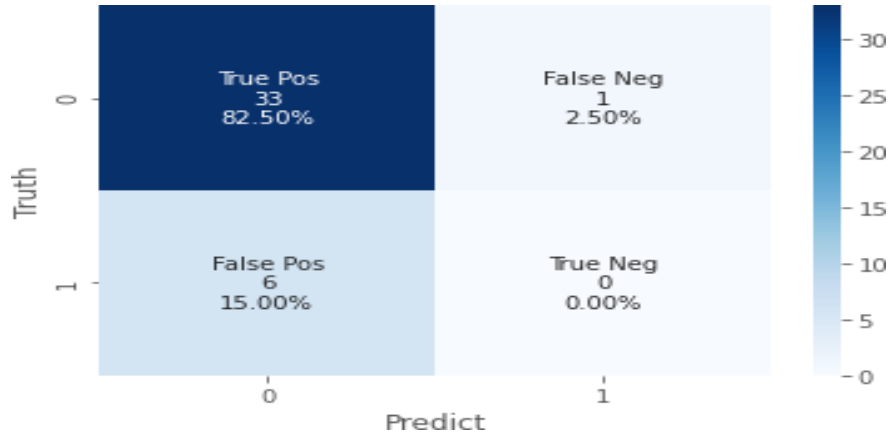


Figure 4.3.12.2: Plotting Support Vector Machine (SVM) Classifier True and False Positive Rate

4.3.13 Plotting Random Forest Classifier:

4.3.13.1 Plotting Random Forest Classification report:

Here we are plotting Random Forest Classification Report. In this plot we are showing how many data are actually true and false detected with respect to train and test data and also show their prediction result.



	precision	recall	f1-score	support
0	0.85	0.97	0.90	34
1	0.00	0.00	0.00	6
accuracy			0.82	40
macro avg	0.42	0.49	0.45	40
weighted avg	0.72	0.82	0.77	40

Figure 4.3.13.1: Plotting Random Forest Classification report

4.3.13.2 Plotting Random Forest Classifier True and False Positive Rate:

Here we are plotting Random Forest Classifier True and False Positive Rate. In this plot we are showing how many data are actually true and false detection rate with respect to train and test data.

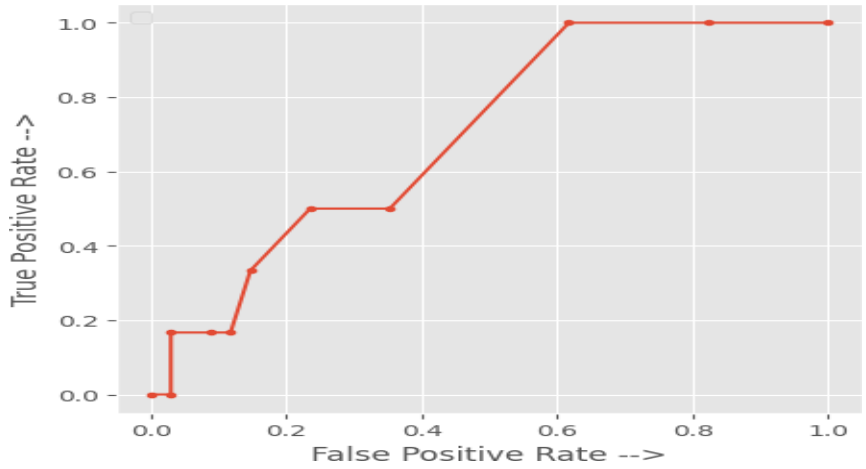


Figure 4.3.13.2: Plotting Random Forest Classifier True and False Positive Rate

4.3.13.3 ROC curve (receiver operating characteristic curve):

Here we are applying ROC curve on our collected dataset according to our four algorithms.

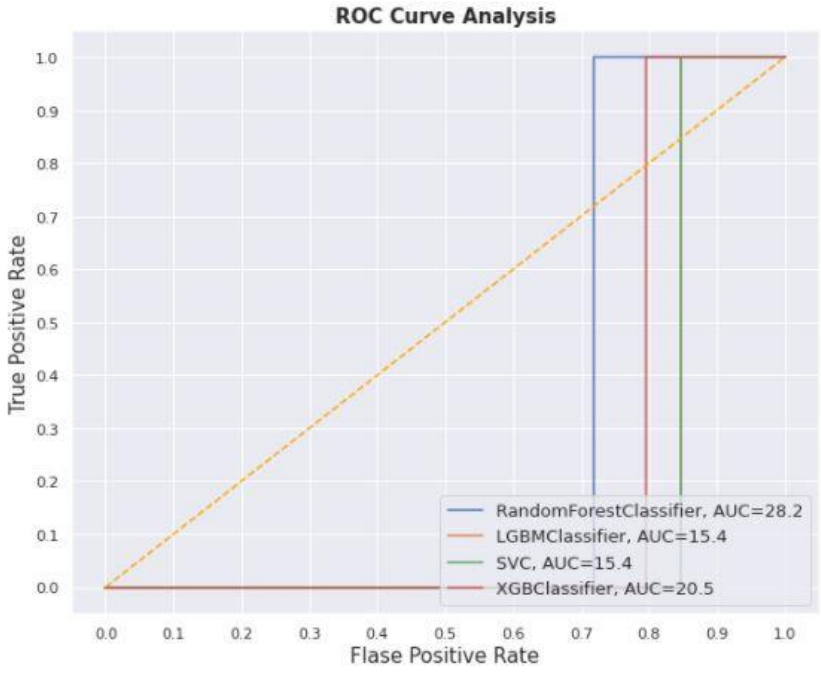


Figure 4.3.13.3: ROC Curve Analysis

4.4 Summary:

In this chapter, we will go over our experimental results analysis. We're displaying a data visualization of our whole system. We exhibit existing data experimental result analysis, our own data experimental result analysis, and lastly merged data experimental result analysis here.We can also predict whose types of people has highly possibility for committing suicide and we also propose some factors which can be followed to prevent suicide.

Gender, age, income, employment and education level are the some of the indicators of suicidal attempt. Whether a person has friends or not, whether he has social fear or not also play major roles in predicting suicidal attempts. Because these leads a human to mental depressed stage, attempt suicide is perfect time for a mentally depressed person.

attempt_suicide	[help] i don't want help	[help] set me up with a date	[help] wingman/wingwoman	[help] date coaching
No	0.363636	0.270793	0.206963	0.158607
Yes	0.392523	0.224299	0.196262	0.186916

We are showing that those people who does not want to help from others they are more suicidal than the others.

attempt_suicide	[improve] none	[improve] other exercise	[improve] joined a gym	[improve] therapy	[improve] meet ups	[improve] join clubs/socual clubs/meet ups	[improve] socual clubs/meet ups	[improve] joined a gym/go to the gym
No	0.373737	0.121212	0.116162	0.080808	0.151515	0.040404	0.090909	0.025253
Yes	0.265306	0.122449	0.061224	0.061224	0.285714	0.081633	0.102041	0.020408

Here, we are seeing how a person improve himself from suicidal attempt. The people who exercised, meet up people, Joins clubs or social meet up, they are attempting suicide than other.

4.5 Prediction Result Accuracy Analysis

4.5.1 Introduction:

In this chapter, we present our prediction result accuracy analysis table in its entirety. First, we display the existing dataset prediction results in the analysis table, then our own dataset prediction results in the analysis table, and finally the merged two datasets prediction results in the analysis table. We also talk about how our society affects us. The methodology for predicting suicide attempts is extremely beneficial to our society. It is capable of detecting and preventing suicide attempts.

4.5.2 Prediction Result Analysis for Existing dataset:

Here, we show Recall, Precision, F1-Score formula.

Recall (R): Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.

Precision (P): Precision quantifies the number of positive class predictions that actually belong to the positive class.

F1-Score (F): F-Measure provides a single score that balances both the concerns of precision and recall in one number.

$$R = \frac{TP}{TP+FN}$$

$$P = \frac{TP}{TP+FP}$$

$$F = \frac{2\times P\times R}{(P+R)}$$

4.5.3 XG Boost Classification prediction report table:

TP	TN	FP	FN
77	2	12	3

	Precision	Recall	f1-score	Support
	0.87	0.96	0.91	80
	0.40	0.14	0.21	14
Accuracy			0.84	94
Macro Avg	0.63	0.55	0.56	94
Weighted Avg	0.80	0.84	0.81	94

4.5.4 LightGBM Classification report table:

TP	TN	FP	FN
75	4	10	5

	Precision	Recall	f1-score	Support
	0.88	0.94	0.91	80
	0.44	0.29	0.35	14
Accuracy			0.84	94
Macro Avg	0.66	0.61	0.63	94
Weighted Avg	0.82	0.84	0.83	94

4.5.5 Support Vector Machine (SVM) Classification report table:

TP	TN	FP	FN
80	1	13	0

	Precision	Recall	f1-score	Support
	0.86	1.00	0.92	80
	1.00	0.07	0.13	14
Accuracy			0.86	94
Macro Avg	0.93	0.54	0.53	94
Weighted Avg	0.88	0.86	0.81	94

4.5.6 Random Forest Classification report table:

TP	TN	FP	FN
77	3	11	3

	Precision	Recall	f1-score	Support
	0.88	0.96	0.92	80
	0.50	0.21	0.30	14
Accuracy			0.85	94
Macro Avg	0.69	0.59	0.61	94
Weighted Avg	0.82	0.85	0.82	94

4.6 Prediction Result Analysis for Our Own dataset:

4.6.1 XG Boost Classification prediction report table:

TP	TN	FP	FN
34	0	6	0

	Precision	Recall	f1-score	Support
	0.85	1.00	0.92	34
	0.00	0.00	0.00	6
Accuracy			0.85	40
Macro Avg	0.42	0.50	0.46	40
Weighted Avg	0.72	0.85	0.78	40

4.6.2 LightGBM Classification report table:

TP	TN	FP	FN
33	0	6	1

	Precision	Recall	f1-score	Support
	0.85	0.97	0.90	34
	0.00	0.00	0.00	6
Accuracy			0.82	40
Macro Avg	0.42	0.49	0.45	40
Weighted Avg	0.72	0.82	0.77	40

4.6.3 Support Vector Machine (SVM) Classification report table:

TP	TN	FP	FN
34	0	6	0

	Precision	Recall	f1-score	Support
	0.85	1.00	0.92	34
	0.00	0.00	0.00	6
Accuracy			0.85	40
Macro Avg	0.42	0.50	0.46	40
Weighted Avg	0.72	0.85	0.78	40

4.6.4 Random Forest Classification report table:

TP	TN	FP	FN
33	0	6	1

	Precision	Recall	f1-score	Support
	0.85	0.97	0.90	34
	0.00	0.00	0.00	6
Accuracy			0.82	40
Macro Avg	0.42	0.49	0.45	40
Weighted Avg	0.72	0.82	0.77	40

4.7 Impact on society:

This artificial intelligence technology can assist us in detecting and preventing suicide attempts. Every day, we read in the news about how many individuals commit suicide around the world. Suicide is becoming one of the leading causes of death in the globe. Suicide attempt prediction is extremely beneficial to our society because many individuals are depressed and we don't know who is attempting suicide. If we can detect this early on, we can prevent it and save a life. It can also be beneficial to medical technicians, doctors, psychiatrists, social workers, and others.

4.8 Summary:

We present our forecast result accuracy analysis in the form of a table. We also talk about how it affects our culture. The methodology for predicting suicide attempts is extremely beneficial to our society.

Chapter-5

Conclusion

5.1 Introduction:

Suicide is a major problem in today's world. If we can detect this early on, we can prevent it and save a life. Suicide is becoming one of the leading causes of death in the globe. We can discern the root cause of many people's depression. Our goal is to gather data and discover risk variables that explain and predict suicide behavior. We are attempting to improve our old prediction approaches in order to use the Machine Learning prediction methodology. Suicide is clearly a major problem in Bangladesh. In Bangladesh, no suicide attempt prediction system is in place. That is why we are attempting to collect data about our country's population and apply some algorithm to assess that data. Also, compare data from other countries. We are teaching a machine to train our own data set in our Artificial Intelligence system. We are leveraging data sets obtained by our friends and family members in our data collection. We are gathering this information in response to some frequently asked questions. In this system, four algorithms are used for comparison and prediction: the Random forest methodology, the XG Boost method, the Support Vector Machine (SVM) methodology, and the Light GBM method. We use another data collection that is not from Bangladesh in this system. We utilize this data set for comparison reasons with our national data set. We are using our suggested approach to identify persons who are at danger of attempting suicide. We are implementing our proposed architecture in our dataset collection. The findings of our proposed system are highly encouraging. This artificial intelligence technology can assist us in detecting and preventing suicide attempts. It can also be beneficial to medical technicians, doctors, psychiatrists, social workers, and others.

5.2 Future Works and Limitations:

In this Suicide attempt prediction system, we use our suggested approach to identify those who are at risk of attempting suicide. We are implementing our proposed architecture in our dataset collection. Our proposed system yields promising results but has several drawbacks. Our biggest drawback is the modest size of our existing dataset, which we obtained ourselves. Another limitation of our collected dataset is people sometime does not give their controversial data. If they gave the real data of their, that would occur a massive change in accuracy level. In the future, we will endeavor to expand our dataset. Our solution outperforms and outperforms the old system in terms of performance and accuracy. However, we believe that the results are insufficient. As a result, we will strive to improve our system's performance and accuracy level in the future.

REFERENCES

1. S. Syed, and I. Amin. "Prediction of Suicide Causes in India using Machine Learning." Journal of Independent Studies and Research-Computing 15, no. 2 (2017): 1-7.
2. K. v. Mens, C. W. M. de Schepper, B. Wijnen, S. J. Koldijk, H. Schnack, P. de Looff, J. Lokkerbol, K. Wetherall, S. Cleare, R. C. O'Connor, D. de Beurs et al. "Predicting future suicidal behaviour in young adults, with different machine learning techniques: a population-based longitudinal study." Journal of affective disorders (2020).
3. V. S. Roglio, E. N. Borges, F. D. Rabelo-da-Ponte, F. Ornell, J. N. Scherer, J. B. Schuch, I. C. Passos, B. Sanvicente-Vieira, R. Grassi-Oliveira, L. v. Diemen, F. Pechansky, F. Henrique, P. Kessler et al. "Prediction of attempted suicide in men and women with crack-cocaine use disorder in Brazil." PloS one 15, no. 5 (2020): e0232242.
4. K. v. Mens, C. W. M. de Schepper, B. Wijnen, S. J. Koldijk, H. Schnack, P. de Looff, J. Lokkerbol, K. Wetherall, S. Cleare, R. C. O'Connor, D. de Beurs et al. "Predicting future suicidal behaviour in young adults, with different machine learning techniques: a population-based longitudinal study." Journal of affective disorders (2020).
5. L. Zheng, O. Wang, S. Hao, C. Ye, M. Liu, M. Xia, A. N. Sabo, L. Markovic, F. Stearns, L. Kanov, K. G. Sylvester, E. Widen, D. B. McElhinney, W. Zhang, J. Liao, X. B. Ling et al. "Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records." Translational psychiatry 10, no. 1 (2020): 1-10.
6. D. D'Hotman, E. Loh, and J. Savulescu. "AI-enabled suicide prediction tools: ethical considerations for medical leaders." BMJ Leader (2020): leader-2020.
7. C. G. Walsh, J. D. Ribeiro, and J. C. Franklin. "Predicting risk of suicide attempts over time through machine learning." Clinical Psychological Science 5, no. 3 (2017): 457-469.
8. R. A. Bernert, A. M. Hilberg, R. Melia, J.P. Kim, N. H. Shah, and F. Abnousi. "Artificial intelligence and suicide prevention: a systematic review of machine learning investigations." International journal of environmental research and public health 17, no. 16 (2020): 5929.
9. A. Biswas, A. Talukder, D. Bhattacharjee, A. Chowdhury, and J. Sanyal. "Machine Learning Based Prediction of Suicide Probability." International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-10 Issue-1, October 2020
10. G. E. Simon, E. Johnson, J. M. Lawrence, R. C. Rossom, B. Ahmedani, F. L. Lynch, A. Beck, B. Waitzfelder, R. Ziebell, R. B. Penfold, S. M. Shortreed et al. "Predicting suicide

attempts and suicide deaths following outpatient visits using electronic health records." American Journal of Psychiatry 175, no. 10 (2018): 951-960.

11. S. M. Bae, S. A. Lee, and S.-H. Lee. "Prediction by data mining, of suicide attempts in Korean adolescents: a national study." Neuropsychiatric disease and treatment 11 (2015): 2367.
12. R. C. Brown, E. Bendig, T. Fischer, A. D. Goldwich, H. Baumeister, and P. L. Plener. "Can acute suicidality be predicted by Instagram data? Results from qualitative and quantitative language analyses." PloS one 14, no. 9 (2019): e0220623.
13. I. M. Harris, S. Beese, and D. Moore. "Predicting future self-harm or suicide in adolescents: a systematic review of risk assessment scales/tools." BMJ open 9, no. 9 (2019): e029311.
14. J. Oh, K. Yun, J.H. Hwang, and J.H. Chae. "Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales." Frontiers in psychiatry 8 (2017): 192.
15. C. R. Glenn, and M. K. Nock. "Improving the short-term prediction of suicidal behavior." American journal of preventive medicine 47, no. 3 (2014): S176-S180.
16. S. Ji, S. Pan, X. Li, E. Cambria, G. Long, and Z. Huang. "Suicidal ideation detection: A review of machine learning methods and applications." IEEE Transactions on Computational Social Systems (2020).
17. M. Sanderson, A. G. Bulloch, J. Wang, K. G. Williams, T. Williamson, and S. B. Patten. "Predicting death by suicide following an emergency department visit for parasuicide with administrative health care system data and machine learning." EClinicalMedicine (2020): 100281.
18. A. Roy, K. Nikolitch, R. McGinn, S. Jinah, W. Klement, and Z. A. Kaminsky. "A machine learning approach predicts future risk to suicidal ideation from social media data." NPJ Digital Medicine 3, no. 1 (2020): 1-12.
19. V. Venek, S. Scherer, L.-P. Morency, J. Pestian et al., “Adolescent suicidal risk assessment in clinician-patient interaction,” IEEE Transactions on Affective Computing, vol. 8, no. 2, pp. 204–215, 2017.
20. M. Denscombe, The Good Research Guide For Small Scale Research Projects. 2010. pp. 14-24, 170-176, 7-8, 330-342, 240-266.