PREDICTION OF SUICIDE ATTEMPTS USING MACHINE LEARNING

BY

Mahmudul Hasan
ID: 191-15-12891
AND
Sayem Ahmed Shifat
ID: 191-15-12353
AND
Pranto Chakrobortty
ID: 191-15-12662

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

Supervised By

Sharmin Akter
Sr. Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Abu Kaisar Mohammad Masum
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

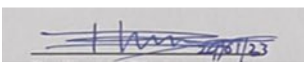
DHAKA, BANGLADESH

JANURARY 2023

APPROVAL

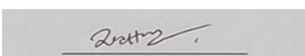
This Research titled **"Prediction of suicide attempts using Machine Learning"**, submitted by **Mahmudul Hasan, ID NO: 191-15-12891, Sayem Ahmed Shifat, ID NO: 191-15-12353 and Pranto Chokrobortty, ID NO: 191-15-12662**. Student ID to the Department of Computer Science and Engineering, Faculty of Science and Information Technology, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc in Computer Science & Engineering and approved as to its style and contents.

BOARD OF EXAMINERS



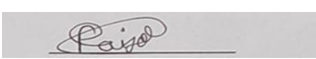
Chairman

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



Internal Examiner

Dr. Md. Zahid Hasan
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Fahad Faisal
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Ahmed Wasif Reza
Associate Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Sharmin Akter, Sr. Lecturer, and Department of CSE Daffodil** International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

____

Sharmin Akter
Sr. Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Kaisar
____

Abu Kaisar Mohammad Masum
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Hasan
____

Mahmudul Hasan
ID: 191-15-12891
Department of CSE

Shifat
____

Sayem Ahmed Shifat
ID: 191-15-12353
Department of CSE

Pranto
____

Pranto Chokrobortty
ID: 191-15-12662
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

At First we would like to express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year thesis successfully.

We really grateful and wish our profound our indebtedness to **Sharmin Akter**, **Sr. Lecturer,** Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest with supportive instructions helped us in the field of deep learning based research, finally we completed our work on "**Prediction of suicide attempts using Machine Learning**". Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan**, Professor and Head, Department of Computer Science and Engineering, Faculty of Science and Information Technology, DIU, for his valuable support and advice to finish our project and also heartiest thanks to other faculty member and the staff of department of CSE, Daffodil International University.

At last, again we want to thank all the good wishers, friends, family, seniors for all the help and inspirations. This research is a result of hard work and all those inspirations and assistance.

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

ABSTRACT

In Bangladesh, there is a rising public health issue around suicide. Targeted suicide control and prevention efforts can be greatly aided by a thorough understanding and forecast of suicide patterns. In this research, after initially analyzing suicide trends and geographical distribution of suicides in Bangladesh, we developed a machine learning model for predicting suicide at the county level for the seven-year period between 2015 and 2022 using publicly available data. Suicide is the intentional act of harming oneself in an effort to end one's life. Suicides frequently have a variety of causes, including despair, financial hardship, mental illness, legal issues, encompassing situations, etc. People's propensity for self-destruction may be a severe problem that is not unique to any one state or nation. Suicide may significantly affect the worldwide death rate, according to data at the global level. Additionally, suicide is among the top twenty causes of death in the globe. Suicide is a subject that has gained increasing cultural attention. Actually, it is one of the major causes of death in the contemporary world. In order to counter this threat, it is crucial to develop precise prediction algorithms based on available data. The research primarily examines suicide data, finds significant risk variables, and forecasts future suicide attempts with a high degree of accuracy. Three machine learning algorithms—logistic regression, random forest, and Naive-Bayes—have been compared with the goal of predicting suicide.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Additionally to the 703 000 suicides per year, many more individuals attempt suicide. Every suicide is a tragedy that profoundly affects the surviving, including the families, neighborhoods, and whole countries. In 2022, the fourth leading cause of mortality globally for those ages 15 to 29 was suicide, which can happen to anybody. Suicide is a worldwide disease that affects all nations, not just those with great incomes. In actuality, suicide rates in low- and middle-income countries increased to above 77% in 2019. Despite being a serious public health concern, suicide may be prevented using quick, reliable, and usually low-cost treatments. The efficacy of national programs depends on a comprehensive multisectoral suicide prevention approach. Predicting and calculating the likelihood of suicide conduct in mental patients is a significant and difficult treatment topic. Suicide attempts have been evaluated and correlated with numerous clinical parameters using a variety of techniques that have been developed and tested. Despite the fact that earlier suicide attempts have long been believed to be a significant predictor of future suicide attempts, just one third of people who attempt suicide have a history of doing so. Despite the widespread use of clinical measures to evaluate suicidal ideation and behavior, these measures' specificity for suicide attempt detection is often inadequate because they commonly mistake high-risk attempters for low-risk individuals. Despite the fact that machine learning algorithms showed great performance that could be evaluated in clinical settings, categorizing and predicting suicide attempts is still challenging. The primary concerns appear to be self-reporting bias of suicide attempts and the low accuracy of any measure used to predict suicide attempts. Inconsistency in predictors is a significant hurdle in addition to these other considerations. When assessing suicide attempts, it is usual to use one or two clinical measures, such as the scores for despondency and suicidal intention. In some situations, this method can demonstrate dependability, but it is easily susceptible to reporting biases on the part of the respondents [1]. When calculating suicide attempts, using different clinical scales simultaneously may help to minimize biases and improve classification accuracy. Life satisfaction, purpose in life, and emotion control are examples of positive psychological qualities that are seldom utilized as factors in evaluating suicide attempts.

Additionally, depressive, anxious, and hopeless psychological states have all been linked to an increased risk of suicide attempt. The level of patients' positivity should be taken into account because suicidality might be caused by a lack of satisfying feelings and emotions [2] the estimation of suicide attempts might be required. Using a specific artificial neural network classifier, we provide a model in this research for classifying suicide attempts in individuals with mental illnesses. In order to create and test this novel model, we assessed the classification power for real suicide attempts of 31 self-report psychiatric and psychological measures, which are routinely used in clinical settings.

1.2 Objective

Since there are many problems in the world that lead to depression, mental sickness, irritability, etc. we have decided to prevent them. First, we'll use social media to identify anyone who has experienced depression or has considered doing so, after which we'll give them some information in the hopes that it will help them avoid it. Finally, we will use our system to call you for a brief interview.

• The gathering of data from social media

• Extraction of data to identify those who have experienced suicidal thoughts.

• Following the prediction, we would work to stop them.

1.3 Motivation

Because psychiatric disorders are known to increase the risk of suicide in the general population and are frequently found to be closely associated with one another, hierarchical clustering could be used to determine the relationships among the potential illnesses that could be causing such a suicidal-risk scenario (concept of syndrome).Additionally, rules might be created using the information collected to validate the connections between the clusters. The ideation-to-action framework's acceptance and the expansion of ideation-to-action theories of suicide are encouraging developments that will significantly increase our understanding of and ability to prevent suicide. However, there are still significant information gaps. These holes restrict our capacity to comprehend suicide and lessen its prevalence and need to be the subject of intense study initiatives in the next years.

1.4 Rational of the Study

This work has a wide range of applications. In order to be more proactive in the support they provide and to provide some unstated context useful to support during the time immediately following a celebrity suicide report, community moderators may make plans for when famous people commit suicide. By doing this, they and other interested/committed volunteers will be able to do both. They will be able to pay particular attention to Redditors who display higher symptoms of distress made worse by the suicide incident since moderators and volunteers will be aware of this behavior on SW. Additionally, people with highly suicidal words and other language constructions may be linked to neighbors who are willing to lend support and assistance. Social support and having more social capital might help people battle such susceptible inclinations.

1.5 Expected Outcome

Suicide cannot be totally prevented, but with better analytical methods, prediction and prevention should be improved. For conventional epidemiological research and medical professionals, predicting the risk of suicide is still a difficult task. This is a result of the complexity of factors that influence suicide and the difficulties in locating a small fraction of individuals who share risk factors among a large population. Finally, our system will try to prevent who is going to spend the worst time and thoroughly think suicide.

- Identification of those who may be considering suicide
- The main objective of the current study was to review prior work and evaluate a novel method for identifying suicide risk.
- This inquiry may yield significant information new knowledge regarding temporal variation in suicide attempt risk and a step toward scalable and effective risk identification.

1.6 Research Questions

We found it really difficult to finish this assignment. To express the ideas and impacts of this issue in order to have a realistic, useful, and accurate response, the researchers would like to offer the following questions.

- Have decades-long studies been done to track the suicidal tendencies of a sample of mentally ill and non-mentally ill patients?
- The functions of evidence-based risk variables in the context of a suicide scenario in a psychiatric population have not yet been examined in research, is that right?
- Has Facebook ever been a platform for social suicide prediction?

1.7 Layout of the Report

The report's broad format is covered in Chapter 1, which also provides as an example of how the project would be presented to the venture's stakeholders, along with its goal, motivation, research questions, and anticipated outcomes.

Chapter two, discusses previous work that has been done in this field. The scope that resulted from their field's constraint is then demonstrated in this second chapter's subsequent portion. Most notably, the key challenges or problems encountered throughout this investigation are discussed.

The theoretical discussion around this research endeavor is described in Chapter 3. In order to address the theoretical component of the research, this chapter elaborates on the statistical methods utilized in the study. This chapter also illustrates the procedural approaches of the machine learning classifier. Confusion matrix analysis is provided in the chapter's last section to approve the presentation and make the precision name of the classifier visible.

The experimental results, performance analysis, and result discussion are all included in chapter 4. This chapter includes a few experimental images to help the project come to life.

Chapter five, discussed along with a summary, recommendations, and conclusions. After the suggestions, the entire project report is anticipated to appear in this chapter. The chapter is brought to a close by pointing out the shortcomings of our efforts, which might ultimately have an impact on other individuals who need to work in this industry.

CHAPTER 2

BACKGROUND

2.1 Introduction

This paragraph will provide pertinent literature, a synopsis of the research, and issues with this study. We will discuss the techniques, correctness, and conclusions of other research publications as they relate to our work under the section on related studies. Under the section on research summaries, we will list all of our related activities. In the difficulties section, we'll go through how we increased accuracy.

2.2 Related Works

In their narrative review published in 2020, D'Hotman et al. [3] come to the conclusion that AI has a significant potential in suicide risk prediction while having ethical reservations regarding the usage of personal data. We did a thorough assessment of the literature, including clinical research employing AI to evaluate suicide risk, to better understand the potential of this technology. This is the first systematic review of the topic that we are aware of. The objective of this review is to evaluate AI's potential for suicide risk prediction and identifying those who are at risk of trying to commit suicide within a population. This technology may help physicians in clinical settings by helping them more precisely identify individuals who are at risk of suicide with the goal of improving suicide prediction skills. To validate this instrument and use it in clinical practice, more research is necessary [4, 5]. Gradus et al. [6] used various Danish databases to analyze data from a relatively large sample of patients in order to predict suicide risk according to gender. They successfully predicted suicide risk using a random forest algorithm (AUC of 0.80 in males and 0.88 in women). Authors Choudhury at all. [7] made sure to include many mechanisms on suicidal difficulties in blogs that discussed celebrity suicides and the progression from mental and emotional problems to suicidal thoughts. Although pricey, questionnaires are a useful instrument for gathering data [8–9]. Suicidal content is a key factor in classifier training. The current dataset has fewer information about suicidal content. Suicidal prevention and identification will be difficult with automated user tweet detection. With a larger dataset and more features, the suggested method looked at a solution that was superior to the earlier study. Compared to surveys or questionnaires, automated programmed detection and identification provides, this can be applied in the actual world. In order to find harmful informational influences in social networks, Okhapkina et al. [10] modified information search techniques and produced a lexicon of terminology for suicidal ideas and actions. Additionally, they created the singular vector decomposition technique and the use of TF-IDF matrices for these matrices. Sawhney et al. [11] enhanced the capability of a random forest classifier to recognize suicidal thoughts in tweets. The logistic regression classification techniques of Aladag et al. [12] were used to identify suicidal content with an accuracy of 80-92%. In order to detect suicidal thoughts utilizing sophisticated DL architectures, neural network models in NLP can outperform conventional ML systems. Sequences can make good use of recurrent neural networks (RNNs) [13]. Relevant data can be preserved using LSTM without the need for external dependencies. According to Sawhney et al. [14], CLSTM-based models are more effective at spotting suicidal ideation than other DL and ML classifiers. Ji et al [15].'s comparison of the value of the different strategies was demonstrated using an LSTM classifier with five ML models. One of the main standards for identifying suicide ideation on social media sites like Twitter and Reddit Suicide Watch was supplied by their study. Many of the risk indicators frequently linked to suicidality in general—as was already mentioned— were not suggestive of any efforts among the ideators who were at high risk. Although despair and depression are also significant predictors of suicidality according to Bolton et al. (2008) [16] and described in Brezo et al. (2006) [17], neither of these factors predicted further attempts after accounting for earlier attempts. According to one theory, these variables may play a part in the development of suicidal thoughts, but less so in the choice of some individuals to act on these ideas and attempt suicide while others choose not to. The degree of melancholy and hopelessness may influence the impulse to commit suicide, but not the actual act of doing so, as would be consistent with Joiner's interpersonal-psychological hypothesis.

2.3 Research Summary

In this paper, an essential and effective method for preventing suicide is the early detection of suicidal thinking. Psychologists used statistical analysis to carry out the great bulk of the study on this subject. On the other hand, deep learning representation learning and feature engineering-based machine learning have been used extensively in computer scientists' research.

If medical experts can see early suicide ideas on microblogging sites like Twitter, They will be able to identify possibly suicidal individuals far more easily, saving countless lives. Deep learning and machine learning techniques may help to enhance the early identification of suicidal ideation and the subsequent early prevention of suicide.

2.4 Challenges

A glossary of phrases was initially created for this purpose by manually annotating anonymized data that was scraped from well-known suicide Web forums. Using search terms that corresponded to the produced lexicon, a dataset of tweets was gathered using the Twitter REST API. In order to train the three machine learning-based baseline models and the three proposed deep learning models, human annotators classified tweets as having suicidal intent or not. In order to detect suicidal thoughts on Twitter as well as other Web forums and Social media, future research may build on this work by studying other deep learning-based architectures.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

In our study, we employed a range of methods, such as decision trees, logistic regression, random forest, and linear regression. On the other side, those are algorithmic predictions based on which we made our decisions. We obtained data from social media, separated it, and then divided the test.

3.2 Research Subject and Instrumentation

Research area that was analyzed and investigated for idea clarification would be the topic of inquiry. Not just for implementation, but also for model development, data gathering, implementation, processing, and model training. In the technology and approach we employed are mentioned in the section titled Instrumentation. We utilized the Windows platform and the python programming language with a variety of packages, including numpy, pandas, skit learn, matplotlib, and seaborn. Google Colab, a free and open-source Python and R programming language distribution for use in information science and artificial intelligence applications, was utilized for the whole training and testing process.

3.3 Workflow

Data gathering, data processing, data resizing and augmentation, model selection, and other workflow phases are included in this study.

Data Collection: By analyzing the raw data we obtained from the website, we developed our own data collection. Since gathering data was so difficult, no dataset in this field is currently accessible.

Data Processing: After being gathered from various sources, all information has been arranged class by class. There are a lot of data with inaccuracies and noise. Prior to implementing the selected dataset in the following phase, we manually process those data.

Model Selection: We select a model in order to train it and validate it for improved accuracy. In order to increase the accuracy of our machine setup, we tried a number of models, and one model was finally picked for the final round of training and testing.

Performance Evaluation: Every result has been described with a graph in this section. We received a few accuracy graphs with validation loss and accuracy after training and testing. Additionally, the confusion matrix was produced, and a table displaying the accuracy, recall, and f1 score measure was created.

Conclusion and Future Work: This section will conclude with a list of next tasks.

3.4 Data Collection Procedure

From Kaggle, we gathered data. These social media platforms—Twitter, Reachout, Weibo, and Reddit—discover people who are depressed about anything, and this dataset is then made available to the public for an experiment. This is how we obtained the 50,000+ above data for prediction and learned which individuals were experiencing the hardest times and had suicidal thoughts. On the other hand, we have gathered data at our university using complicated forms after receiving their input.

3.5 Data Processing

There is information planning in the data handling system. The outcome of our negotiations with push information often depends on the information that has been preprocessed. The information will be pre-processed more efficiently if it is, and the outcome will be more accurate. It is, in a single sentence, the opening obstacle for this type of inquiry-based activity.

3.5.1 **Data Preparation**

Following data collection, it was imported into Google Colab for organization and cleanup before being scaled for more minimal weight, which resulted in a more minimal operating time so that is suitable further processing and analysis. After that machine learning (ML) algorithms are exploring and visualizing the data.

- Loading the data
- Familiarizing with data
- Visualizing the data
- Data Preprocessing & EDA
- Splitting the data
- Training the data
- Model Performance Comparison
- Statistical Tests

3.6 Proposed Methodology

Many people's attention has been drawn to suicide detection. Researchers are concerned about the recent rise in suicide years and has received significant research from numerous perspectives. The investigational methods used. Suicide also encompasses various disciplines and approaches, such as clinical procedures involving patient-clinician interaction automated detection using content produced by users. There are numerous machine learning techniques used to trigger automatic detection. On the expertise and direct contact of the clinician. Clinical interviews and suicide risk assessment measures can provide useful hints for predicting suicide. did a survey and interview study using Weibo, a Chinese version of Twitter, to find out how engaged suicide attempters are with direct messaging intervention. Social media posts by users on websites disclose a wealth of information, including their preferred languages. Using exploratory data analysis on user-generated content, it is possible to gain insight into the linguistic habits and suicide attempters' language use. The thorough study includes topic modeling inside posts on suicide, statistical linguistic aspects, and lexicon-based filtering. The tweets gathered may have suicidal ideas, however only tweets that use the word "suicide" are regarded to have suicidal ideation. The tweets are then manually gathered depending on rules for annotations. The total number of tweets included in the collection, then, is 60,188, including both good and negative tweets.

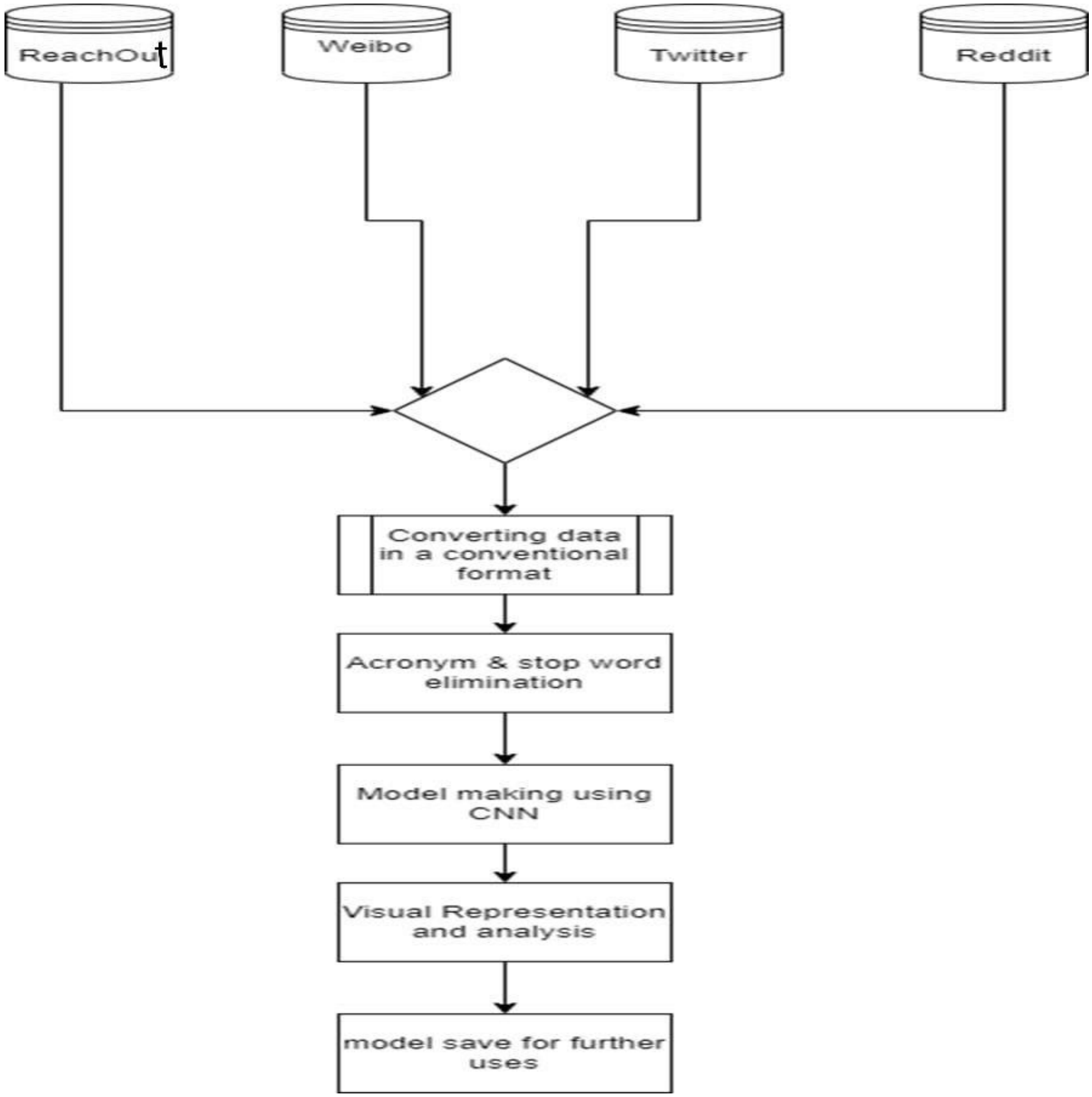


Fig 3.6.1: Total workflow

The provided train data is used to train the Random Forest, Naive Bayes, Linear Regression, and Logistic Regression models. We conducted our testing processes once we had finished our instruction.

3.6.1: Random forests

To generate random forests, sometimes referred to as random decision forests, an ensemble learning approach for classification, regression, and other problems, several decision trees are formed during training. In a classification exercise, the class that the majority of the trees select as their output is known as the random forest. Regression tasks are given with the mean or average forecast of the several trees. Decision trees commonly overfit their training sets, therefore this issue is addressed by using random decision forests. Random forests normally perform better than decision trees, although random forests are less accurate than gradient enhanced trees. The performance of the data, however, could be affected.

3.6.2: Naïve bayes

A straightforward learning technique called Naive Bayes makes the strong assumption that the characteristics are conditionally independent given the class and applies Bayes' rule. Despite the fact that naive Bayes frequently produces competitive classification accuracy, this independence assumption is frequently broken in reality. Naive Bayes is frequently used in practice as a result of this, along with its computational efficiency and other other desirable properties.

3.6.3: Linear Regression

When we learn a linear regression model, we utilize the data that are currently available to estimate the values of the coefficients that are used in the representation. Here, we'll quickly review four methods for creating a linear regression model. This is enough knowledge to understand the calculations and trade-offs but not enough to implement them from scratch. Given how well the model has been explored, there are many more ways. Ordinary Least Squares should be noted since it is the technique that is generally employed the most. Gradient Descent should also be noted because it is the method that is most frequently taught in machine learning classrooms. We got the prediction evaluation.

Fig 3.6.3: Prediction evaluation by Linear Regression

3.6.4 Logistic Regression

In statistics, the logistic model, sometimes called the logit model, employs the log-odds of the event as a linear combination of one or more independent variables to describe the likelihood that the event will occur. In regression analysis, a logistic model's parameters are estimated using logit or logistic regression (the coefficients in the linear combination). The independent variables can each be either binary variables (two classes, denoted by an indicator variable) or continuous variables. Formally, binary logistic regression has a single binary dependent variable with two values denoted by the letters "0" and "1," whereas the independent variables can each be either continuous variables or binary variables. Finally, we got the prediction evaluation using this model.

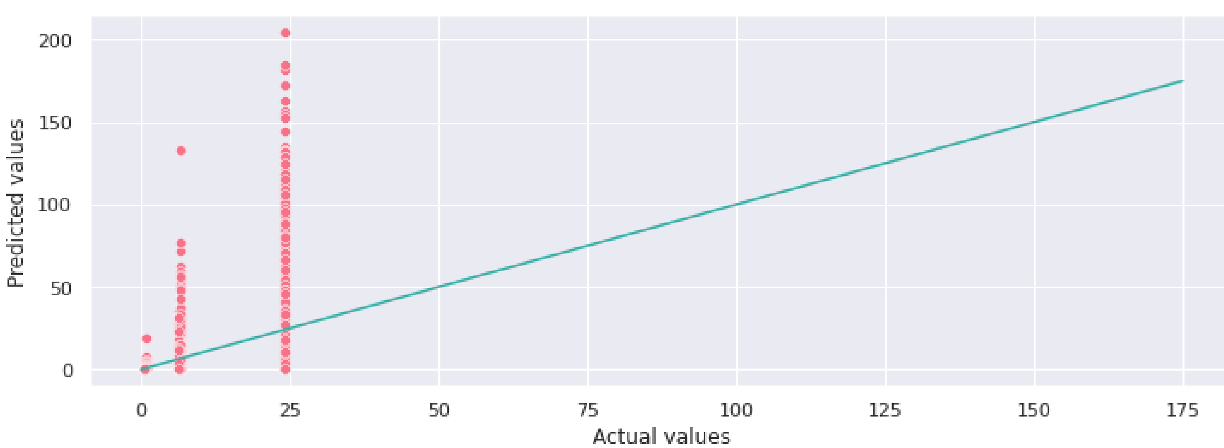


Fig 3.6.4: Prediction evaluation by Random Forest

3.6.5 Decision Tree Regression

In the shape of a tree structure, decision trees construct regression or classification models. It progressively develops a decision tree to go along with the breakdown of a dataset into smaller and smaller subgroups. A tree containing decision nodes and leaf nodes is the end result. Two or more branches, each indicating a value for the characteristic being checked, can be found on a decision node. An assessment of the numerical target is represented by a leaf node. A decision tree's root node is the topmost decision node and the best predictor. Numerical and categorical data may both be handled by decision trees. Finally, we got the prediction evaluation using this model.

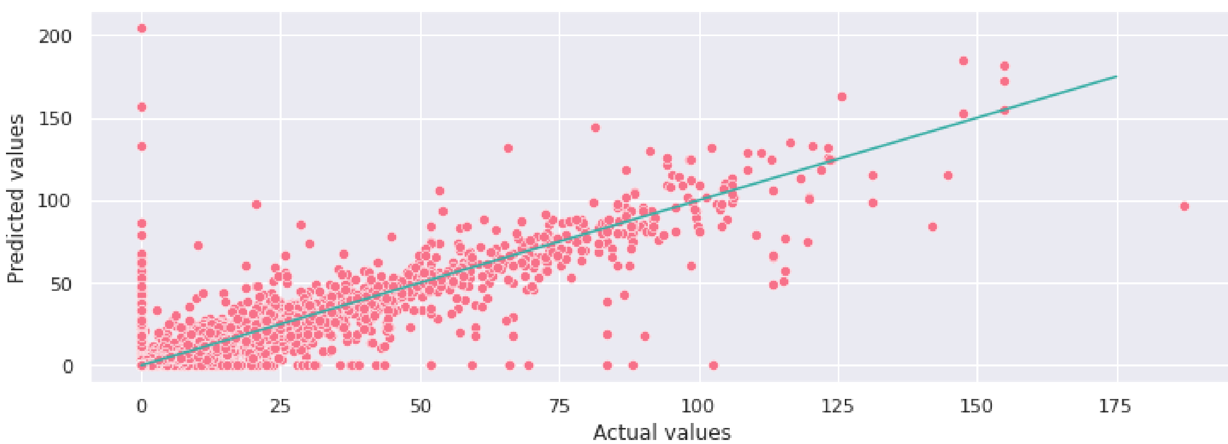


Fig 3.6.5: Prediction evaluation by Decision Tree

3.7 Training the Model

After collecting data we distributed the dataset and data processing involves many different procedures and techniques then loading the data, Familiarizing with data, Visualizing the data, Data Preprocessing & EDA, Splitting the data, Training the data, Model Performance Comparison, and Statistical Tests.

3.8 Implementation Requirements

A list of prerequisites for such an image classification task has been created after a thorough study of all relevant statistical or theoretical ideas and procedures.

The following are likely essential items:

- ➢ Hardware and Software Prerequisites
  - Operating System (Windows 10 or above)
  - Hard Disk (minimum 500-600 GB)
  - Ram (Minimum 4-6 GB)
- ➢ Developing Tools
  - Python Environment
  - Google Colab

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

We obtained the suicide rates from the prediction that we employed a variety of models, such as Random Forest, Nave Bayes, Decision Tree, Linear Regression, and Logistic Regression, after collecting the information. These models gave us the overall quantity of suicidal level, which we then assessed for prevention. They also provided precision, recall, and accuracy.

4.2 Performance Evaluation

The outcomes of the proposed method's evaluation in terms of metrics are shown in [Table 4.3.1] with an emphasis on the negative category as opposed to the positive and neutral categories. This displays the data and results of the feature extraction. All classifiers employ the term frequency value of each word to show the relevance of features, but the contribution of the features used has given various outcomes. Following sentiment analysis, various classifiers are employed to predict the suicidal level from the retrieved characteristics. NB and LR have equal precision and recall metrics, whereas the various features in NB show less fluctuation in accuracy. While NB and RF classifiers perform similarly in terms of precision and F1 scores, RF approach outperformed LR in terms of accuracy by taking into account all available information.

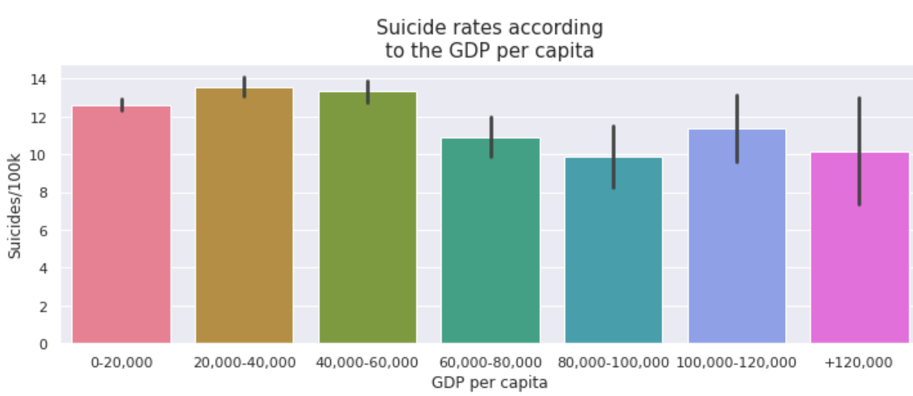


Fig 4.2.1: Suicide rates according to the GDP per capita

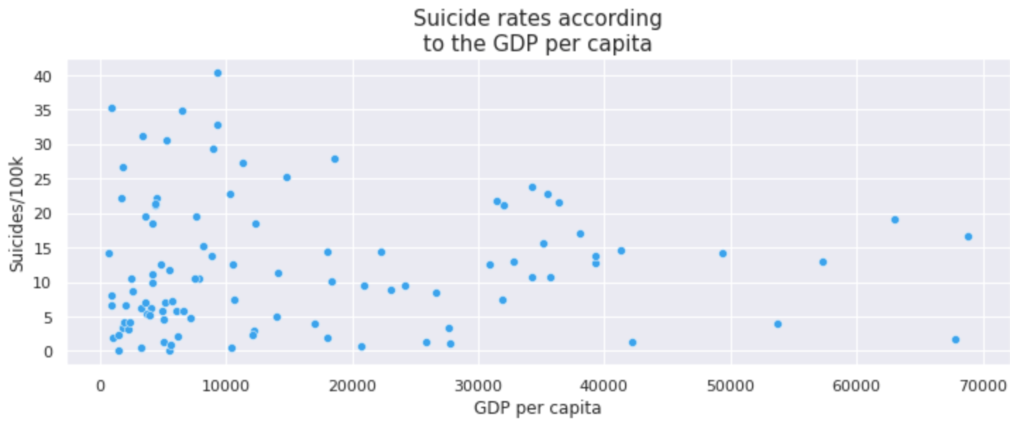


Fig 4.2.2: Suicide rates According to the GDP per capita

4.3 Result Discussion

As a result, the suggested model's accuracy was 0.99% with 0.98% precision, 0.94% recall, and 0.99% score using features and classifier approaches on trained datasets.

Table 4.3.1: Accuracy

Model	Category	Precision	Recall	F-score	Accuracy
Random Forest	**Positive**	**0.56**	**0.51**	**0.47**	**0.99**
	Negative	**0.95**	**0.97**	**0.99**	
Naïve Bayes	**Positive**	**0.27**	**0.23**	**0.25**	**0.97**
	Negative	**0.94**	**0.98**	**0.99**	
LinearRegression	**Positive**	**0.31**	**0.29**	**0.31**	**0.91**
	Negative	**0.92**	**0.94**	**0.98**	
Logistic Regression	**Positive**	**0.29**	**0.31**	**0.29**	**0.85**
	Negative	**0.98**	**0.93**	**0.95**	
DecisionTreeRegression	**Positive**	**0.35**	**0.37**	**0.29**	**0.96**
	Negative	**0.97**	**0.91**	**93**	

CHAPTER 5

CONCLUSION, RECOMMENDATION AND FUTURE WORKS

5.1 Overview of the Study

Finally, we've already covered all the ways we obtained data from Twitter, Reddit, Weibo, and Reachout in this essay. After gathering all of this data, we trained five different types of models to determine their precision, recall, F score, and accuracy. Once the preventative measures have been chosen by the prediction model.

5.2 Conclusion

In this essay, the topic of suicidal identification is investigated using tweets from Twitter, Reddit, Weibo and Reachout. The tweets are acquired automatically and inexpensively. By obtaining and analyzing the unexplored data from the Twitter API, it provides information that can enhance our understanding of suicide behavior and ideation. Though the detection of suicidal ideation can advance with the use of complicated models, other aspects or traits, such as historical data, can be the future direction.

5.3 Future Works

Our main objectives are to prevent people from being depressed, mentally ill, or feeling depressed in general.

- Forecasting the suicide rate.
- Protection against mental injury.
- We'll give them some information that might deter them.
- Suicide risk assessment tools and clinical interviews can offer helpful cues for suicide prediction.

REFERENCES

[1] Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull.. 2017;143(2):187–232.

[2] Le Glaz A, Haralambous Y, Kim-Dufor D-H, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res. 2021;23(5):e15708.

[3] D'Hotman D, Loh E. AI enabled suicide prediction tools: a qualitative narrative review. BMJ Health Care Inform. 2020;27(3):e100175

[4] Torous J, Larsen ME, Depp C, Cosco TD, Barnett I, Nock MK, et al. Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: a review of current progress and next steps. Curr Psychiatry Rep. 2018;20(7):51.

[5] Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim H-C, et al. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep. 2019;21(11):116.

[6] Gradus JL, Rosellini AJ, Horváth-Puhó E, Street AE, Galatzer-Levy I, Jiang T, et al. Prediction of sex-specific suicide risk using machine learning and single-payer health care registry data from Denmark. JAMA Psychiatry. 2019;77(1):25.

[7] M.D. Choudhury, M. Kumar, M. Dredze, and G. Coppersmith, Detecting changes in suicide content manifested in social media following celebrity suicides, HT ACM Conference Hypertext Social Media, pp. 85-94; 2015.

[8] P. Burnap, and J. Scourfield, Analysing the connectivity and communication on suicidal users using Twitter, Computer Communication, vol. 73, pp. 291-300; 2016

[9] S.R. Nayak, and J. Mishra, A Modified Triangle Box-Counting with precision in Error fit, Journal of Information and Optimization Sciences, vol. 39, pp. 113-128; 2018.

[10]. Sawhney, R.; Manchanda, P.; Singh, R.; Aggarwal, S. A computational approach to feature extraction for identification of suicidal ideation in tweets. In Proceedings of the ACL 2018, Student Research Workshop, Melbourne, Australia, 15–20 July 2018; pp. 91–98.

[11]. Alkahtani, H.; Aldhyani, T.H.H. Artificial Intelligence Algorithms for Malware Detection in Android-Operated Mobile Devic-es. Sensors 2022, 22, 2268. [CrossRef]

[12]. Wang, C.; Jiang, F.; Yang, H. A hybrid framework for text modeling with convolutional rnn. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; ACM: New York, NY, USA, 2017; pp. 2061–2069.

[13]. Sawhney, R.; Manchanda, P.; Mathur, P.; Shah, R.; Singh, R. Exploring and learning suicidal ideation connotations on social media with deep learning. In Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Brussels, Belgium, 31 October–1 November 2018; pp. 167–175.

[14]. Ji, S.; Yu, C.P.; Fung, S.-F.; Pan, S.; Long, G. Supervised learning for suicidal ideation detection in online user content. Complexity 2018, 2018, 6157249. [CrossRef]

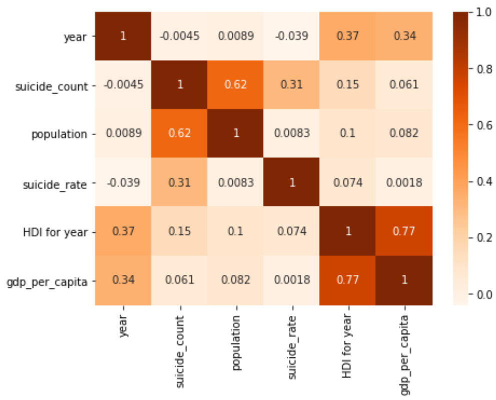
[15]. Bird, S.; Klein, E.; Loper, E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2009.

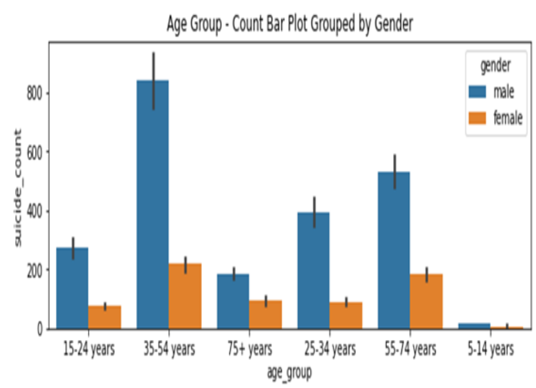
[16]. Bolton JM, Cox BJ, Afifi TO, Enns MW, Bienvenu OJ, Sareen J. Anxiety disorders and risk for suicide attempts: findings from the Baltimore Epidemiologic Catchment Area follow-up study. Depression and Anxiety 2008;0:1e5

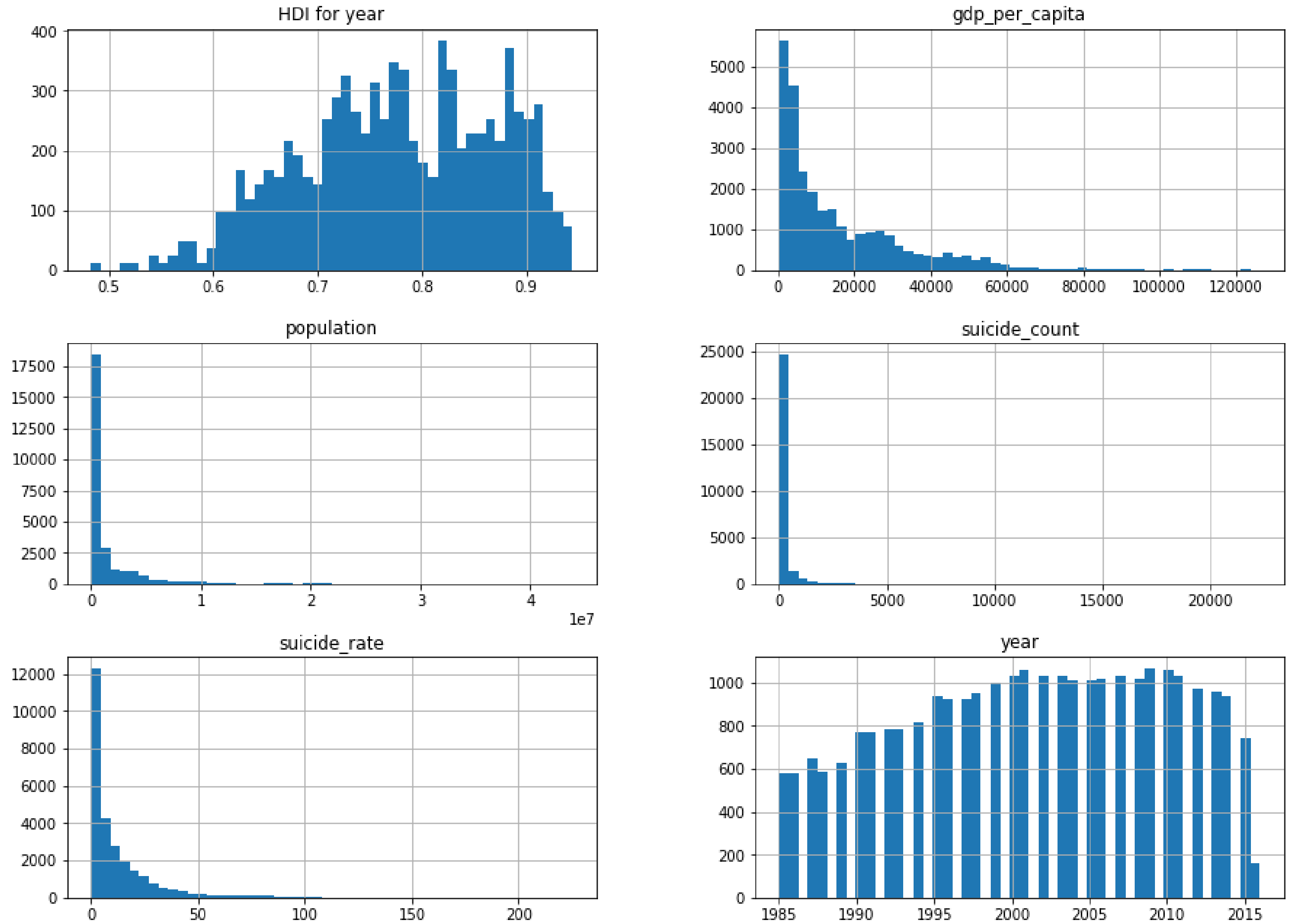
[17]. Brezo J, Paris J, Turecki G. Personality traits as correlates of suicidal ideation, suicide attempts, and suicide completions: a systematic review. Acta Psychiatrica Scandinavica 2006;113:180e206.

APPENDIX

Our society now has a desperate need to lower the suicide rate. Our contribution is intended to aid the community in early detection of suicidal thoughts. In order to identify suicidal thoughts, we have examined text using conventional machine learning. Later, we demonstrated how isolating the top characteristics decreases complexity and boosts accuracy. By choosing features in the future based on a better comprehension of the sequential link between phrases, deep learning approaches can significantly enhance accuracy. We determine the suicide rate by gender and calculate GDP per capita after applying machine learning.






HDI for year
0 100 200 300 400
0.5 0.6 0.7 0.8 0.9
gdp_per_capita
0 1000 2000 3000 4000 5000
0 20000 40000 60000 80000 100000 120000
population
0 2500 5000 7500 10000 12500 15000 17500
0 1 2 3 4
1e7
suicide_count
0 5000 10000 15000 20000 25000
0 5000 10000 15000 20000
suicide_rate
0 2000 4000 6000 8000 10000 12000
0 50 100 150 200
year
0 200 400 600 800 1000
1985 1990 1995 2000 2005 2010 2015

PLAGIARISM REPORT

Report

ORIGINALITY REPORT

25%	19%	16%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Theyazn H. H. Aldhyani, Saleh Nagi Alsubari, Ali Saleh Alshebami, Hasan Alkahtani, Zeyad A. T. Ahmed. "Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models", International Journal of Environmental Research and Public Health, 2022 Publication	6%
2	www.tandfonline.com Internet Source	2%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
4	discovery.researcher.life Internet Source	2%
5	www.frontiersin.org Internet Source	2%
6	www2.psych.ubc.ca Internet Source	2%
7	Submitted to University of Exeter Student Paper	1%