SENTIMENT ANALYSIS USING DIFFERENT MACHINE LEARNING TECHNIQUES ON SOCIAL MEDIA FOR DETECTING SUICIDAL TENDENCY.

BY

NAME: SHAIFUL ISLAM SHAWON
ID: 191-15-2572

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

Name: Ms. Tania Khatun
Designation**: Assistant Professor**
Department of CSE
Daffodil International University

Co-Supervised By

Name: Mr. Md. Mahfujur Rahman
Designation**: Sr. Lecturer**
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY,2024

APPROVAL

This Project titled "**SENTIMENT ANALYSIS USING DIFFERENT MACHINE LEARNING TECHNIQUES ON SOCIAL MEDIA FOR DETECTING SUICIDAL TENDENCY**", submitted by **SHAIFUL ISLAM SHAWON** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *15 July 2024*.

BOARD OF EXAMINERS

Dr. Md. Taimur Ahad (MTA) — **Board Chairman**
Associate Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Shayla Sharmin

Ms. Shayla Sharmin (SS) — **Internal Examiner 1**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Mayen Uddin Mojumdar (MUM) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

MIslam

Dr. Md. Manowarul Islam — **External Examiner**
Associate Professor
Department of CSE
Jagannath University

DECLARATION

I hereby declare that, this project has been done by us under the supervision of, **Ms. Tania Khatun, Assistant Professor, Department of CSE** Daffodil International University. I 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:

Shayla Sharmin 30.9.24

Ms. Tania Khatun
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

for Shayla Sharmin 30.9.24

Mr. Md. Mahfujur Rahman
Sr. Lecturer
Department of CSE
Daffodil International University

Submitted by:

30-09-24

Shaiful Islam Shawon
ID: 191-15-2572
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

I really grateful and wish my profound my indebtedness to **Ms. Tania Khatun**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*machine learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express our heartiest gratitude to Supervisor, Co-Supervisor, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

Suicide is a major problem in latest civilization. Previously diagnosis and reduction of suicide tries are crucial for saving people's lives. Suicidal thoughts can be detected using clinical approaches involving social workers or specialists, as well as ML methods using DL to identify online social material. This study is the first to introduce and describe the approaches in these areas. In this regard, the issue of creating ways for detecting persons who are predisposed to suicide conduct is becoming increasingly relevant. One of the approaches of this research is to look for typological aspects of suicide-related speech utilizing mathematical linguistics, computerized text processing, and machine learning. In foreign science, texts written by persons who were motivated by suicide are researched using automatic text processing methods, machine learning approaches, and models designed to classify whether the content is connected to suicide or not. It should go without saying that analyzing suicide notes together with other writings written by individuals who have taken their own lives is essential to developing techniques for detecting individuals who are at risk of suicide. There are two datasets such as suicidal and sentimental with total 2520 data and total 16 attributes. Passive-Aggressive, Random Forest, logistic regression, SVM, MultinomialNB algorithms are using for sentiment dataset which find best Passive-Aggressive (73%) accuracy and XGBClassifier (96%) accuracy for suicidal datasets.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

The WHO reports that about 800,000 individuals are killed by suicide each year, i.e. Every 40 seconds, someone commits suicide, yet only 30% of those who do so have previously disclosed their intentions, based on the data that is now available [1]. Consequently, it is imperative to devise strategies targeted at detecting individuals who have suicidal tendencies and averting their attempted suicide. Information flows from social networks are large and dynamic, producing substantial amounts of information. Social networks' capacity is continuously growing, which also leads to a rise in the speed at which data is transmitted and processed. Based on the use of contemporary big data analysis technology, information social networks, which at first glance seem to be scattered, can be disseminated based on an enormous array of factors, including those that are specific to individual users as well as those that are shared by particular user groups.

This study examines supervised and unsupervised learning algorithms to detect suicide thoughts in youth and adolescents in Kazakhstan using the well-known social network "Vkontakte". Second only to Facebook in terms of user base size in Europe is VK (formerly known as VKontakte) [2]. Suicidal ideation is understood to be an inclination to take one's own life, including everything from melancholy to a detailed plan for attempting suicide to a severe obsession with self-destruction [3]. Suicide ideators (also known as planners) and attempters (also known as completers) are identifiable as at-risk persons [4]. In research communities, there is much debate over the link between these two groups. Most people who consider suicide don't actually try to end their lives, according to some research. Klonsky and colleagues [5] have posited that most of often mentioned danger elements connected to suicide are really indicators of thoughts of suicide as opposed to the actual path from concept to try. A suicide attempter and planner, however, might resemble "several variables assumed to be risk factors for suicidal behavior," according to Pompili et al. [6]. With the shared goal of lowering 10% reduction in suicide rates by 2020, initially As a national suicide prevention approach aimed at the global market, the identification of suicidal thoughts has been developed and implemented in WHO member states [7]. The outcomes suggest that the system can correctly forecast suicide cases. The achievements of this work are summarized as follows:

- In order to achieve sentiment analysis utilizing several machine learning methods on social media for detecting suicidal tendency, A machine learning framework that is pre-trained using models is suggested.
- A comparative study of the studies conducted by various researchers
- Finding each model's optimal performance and then testing it again regarding the test planned to make sure the model is more broadly applicable.

In Section 2, the techniques employed for this study are explained. The quantitative findings and discussion are provided in part 3, and the methodology for comparing this study to earlier ones is demonstrated in chapter 4. The study's result is eventually revealed in chapter 5.

1.2 Motivation

Most young people use social media, but in recent years, it has become a powerful "window" into the mental health and overall wellness of its users. Through anonymous participation in numerous online communities, it offers a platform for public conversation on socially taboo themes. Over 20% of people who attempt suicide and 50% of people who accomplish suicide frequently leave suicide notes [8]. Therefore, any written indication of suicide is considered concerning, and the person should be questioned about their personal feelings. Social media material, including tweets, blog entries, online notes, and forum messages, is typically captured in the moment and is highly maintained, claim Choudhury et al. [9]. The WHO believes that depression may be a widespread global illness that impacts a vast number of individuals of all ages. Many problems, such as a lack of qualified professionals, societal stigma, incorrect diagnosis, and so forth, impede the detection and treatment of periodontal disease. The study undertaken in this work requires the text knowledge that Twitter is the selected knowledge source, where people tweet about their tales, ideas, feelings, hopes, and mental states. Our study's goals are listed here:

- To ascertain the result of thoughts of suicide.
- Better life.
- To look at the linked problem

1.3 Rationale of the Study

The traditional clinical psychology research employs a labor-intensive theoretical method to identify suicidal tendencies. This labor-intensive, subjective engineering comes after the drawn-out in-person conversation. The main problem with typical Sessions for counseling is that 80 percent of those risk feel uncomfortable talking about their potential levels of stress and anxiety [10]. Given this, it becomes critical to identify those who pose a danger in order to reduce the suicide rate, which is thought to have been avoidable [11]. In view of this, research on the quantification of suicidal propensity may enhance national government by revealing the social media hotspots for suicidal thoughts. Within the medical field, it's critical to forecast the resources that will be needed in the near future to support medical intervention for the diagnosis, treatment, and management of mental illness. People at risk can be identified more quickly and affordably thanks to these automated mental health forecasts. Suicide risk detection models are utilized by social media sites such as Facebook to identify potentially at-risk individuals and provide them with assistance [12]. Researchers have discovered that there is a high correlation between social media use and emotional expression by users [13], and that around 80% of people report their suicide thoughts on social media [14]. Suicidal risk assessments have benefited from the social media mental health prediction [15]. By identifying new study topics from this paper, an exhaustive literature analysis on predicting suicidal inclination may allow academic scholars to enhance this field. This gives us motivation to deal with these issues.

1.4 Research Question

- What is this database's most important features?
- How does this research's algorithm function?
- How can suicidal tendency will be detected early?
- If a person develops mental health what is the success rate?

1.5 Output

Our goal is to identify suicidal tendencies. With the help of our initiative, we will be able to determine whether mental health is impacted and, if so, what proportion of it is affected. We have developed this program with an eye on the future, hoping to identify suicidal tendencies at an early stage and provide effective therapy. In the detection experiment, we used a programming language known as machine learning. Two different types of datasets were used for this study; one included 13 characteristics, while the other had just two. We also employed an alternative algorithm. Thus, we anticipate that our approach will reliably and accurately detect propensity.

1.6 Project Management and Finance

Project financial supervision, often known as accounting, is the oversight of the budget of a project element, including its profit, revenue, cost. It blends billing, estimate, finance, budgeting, and project expenditure management to achieve this. Healthcare executives of today are always improving and fine-tuning their processes in order to save money, boost customer satisfaction, and improve patient care. Over the past ten years, there has been a notable growth in the healthcare industry's use of project management approaches. Because these concepts help with cost management, risk reduction, and overall project results, they have become more and more important for facilities. In the healthcare industry, there are several electronic framework implementations. Because of this, project management has become one of the most in-demand corporate competencies. Project financial management, often known as project accounting, is the administration of elements, including its cost, profit and income. It blends billing, budgeting, finance, project expenditure and estimate management to achieve this. Among the many components of the executive's financial duty, preparing a viable venture is undoubtedly the most important. The next task is to manage that budget throughout the project in order to make sure that the work is completed within the budget that has been allocated. Information gathered from the study of the roles and duties in project management was gathered in order to better understand evolving trends of roles and responsibilities. Meeting planning and scheduling, overseeing the investigation as a whole, and data entry into databases are all included in project management. The timely dissemination of meeting agendas and minutes detailing the study's progress and action items served as a record of that progress. Over time, these materials gathered to form an archive that can be accessed centrally through the Madcaps Interaction Network. In order to provide superior quality information can impact policy needs, management, illness avoidance and Performing health-related research that requires consideration of regional and national political and social contexts requires proficient project management. This happens because effective project management makes it possible to administer and adhere to the right tools, techniques, strategies, and methods in a predictable manner. The population's health may be improved by integrating project management skills into initiatives to set up health care systems in nations with poor and moderate incomes. The schematic of research management and finance using Figure 1.1 may be found below:

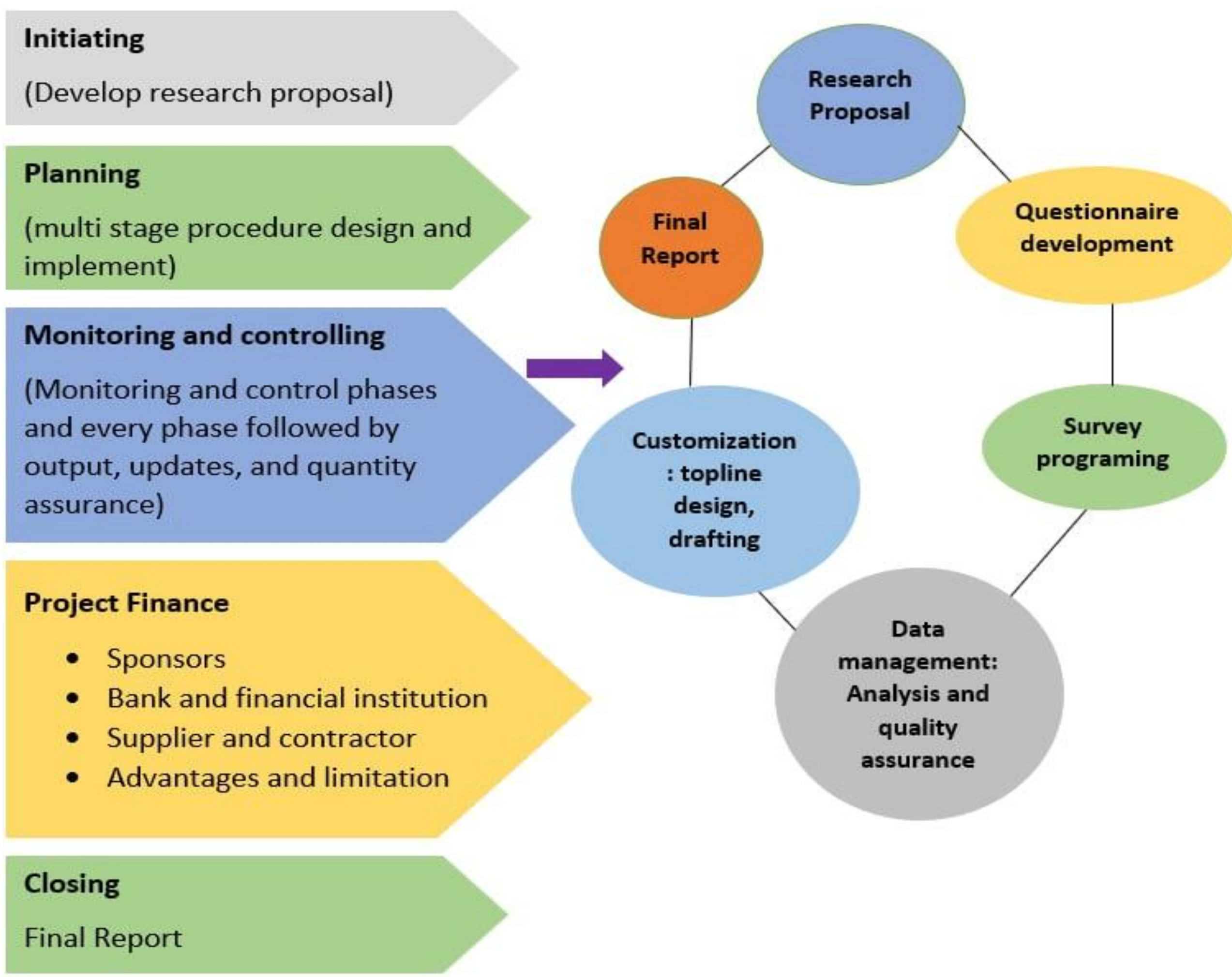

Figure 1.1: Research management and finance

1.7 Research Layout

- In chapter 1 says the ideas of " Sentiment analysis using different machine learning techniques on social media for detecting suicidal tendency". Aside from the objectives, goals, and results of our investigation.
- The challenge, results, and a brief synopsis from earlier relevant research projects are covered in the second chapter part.
- In Chapter 3, the research approach is discussed.
- Chapter 4 goes into detail on the experimental results.
- Our societal influence on environmental consequences, etc., is covered in Chapter 5.
- The research's summary, conclusion, recommendation, and implications are covered in Chapter-6.

CHAPTER 2

Background study

2.1 Preliminaries

The intentional injuring of oneself to terminate one's own life is known as suicide, and as a result, they pass away. Worldwide and in India, one of the leading causes of death is suicide. The multisite intervention study on suicidal behaviors was initiated by (WHO) in 2000 with the objective of expanding Understanding suicidal behaviors and the efficacy of modifications for individuals who attempt suicide in culturally varied populations settings globally [16]. India has a suicide rate of 16.5 per 100,000 persons. Thailand is ranked third in the area with a suicide rate of 14.4, whereas Sri Lanka is ranked second with a rate of 14.4. In India alone, 10.5 out of every 100,000 persons commit suicide annually [17]. This sparked a wide range of conversations on how to determine whether someone has a mental illness and may try suicide in order to give the appropriate counseling and therapy to address the problem. It is common knowledge that people, or more precisely, Internet users, spend the majority a result of the accessibility of the Internet. The user has developed a virtual world addiction as a result of chatting, exchanging photos, posting status updates, leaving comments on other people's posts, etc. social networking sites with chat rooms and blogs. In order for many individuals and organizations to interact and exchange ideas, opinions, and video sites, information, websites, and social media platforms are now necessary. [18]. Social media is widely used, which generates a lot of data that may be used for study, mood detection, etc.

2.2 Related Works

There have been a lot of studies conducted in recent years that show the influence of social media on suicidal ideation.

Drus [19] presents the findings of a review that examined the techniques, social media platform, and applications of sentiment analysis in social media. User-posted text, audio, video, and photo content makes up a sizable portion of the raw data on social media. Sentiment analysis is a useful instrument for converting data into insightful agreement. Utilizing the dependable and trustworthy databases— Scopus, ACM, IEEE Xplore, Emerald Insight, Science Direct, — A comprehensive analysis of studies released between 2014 and 2019 was carried out. 24 out of 77 articles were selected for further examination after the first thorough screening of the paper. Based on the study's objectives, the publications have been reviewed. Considering the goals of the study, the publications have been reviewed. The findings indicate that opinion-lexicon analysis was used in the majority of articles to evaluate the mood of texts on the internet. Data was gathered from websites that mostly Twitter, microblog, and sentiment analysis applications were seen in the fields of politics, business, global events and healthcare.

Birjali [20] put us in a trying situation and discuss feelings that could lead to suicidal thoughts. They specifically suggest creating a suicide-related language in order to overcome the dearth of terminological materials on the subject. They then suggest looking at Weka as a tool based on ML techniques may portion valuable information from Twitter data gathered by Twitter4J in order to conduct a more thorough study. As a result, a WordNet-based method for performing study of semantic differences between training and data set tweets is suggested. The findings of our experiments show that our approach, which uses Twitter Data to extract predictions of suicidal ideas, is according to ML algorithms and semantic sentiment analysis.

Ji, S. [21] approaches from these categories are introduced and discussed in detail for the first time in this survey. Suicidal ideation detection domain-specific apps are examined based on online user content, surveys, suicide notes, electronic health records, and further information sources. To help with future study, a number of particular tasks and datasets are provided and summarized. Lastly, we offer an overview of the present work's shortcomings and suggest future lines of inquiry.

Abdulsalam [22] examines current approaches that employ machine learning techniques to identify social media suicide ideas users. The viability and efficacy of employing social media sites like Weibo, Twitter and Reddit for the identification of suicide thoughts have been extensively studied. Less research has been done on Arabic; The majority of research has focused on depression and suicide detection techniques for widely spoken languages, such as English. Research is thus required for the identification of suicide thoughts in Arabic, given the increasing amount of Arab social media users world.

Dandekar [23] were several research problems listed in order to offer an exhaustive picture of their work. The algorithms are intended to examine the tweet in order to identify emotions and identify instances of suicide ideation among users of social media. Without verifying the accuracy of the tweets, the algorithm analyzes them to forecast sadness. The primary need for a model is its ability to predict outcomes exactly, since many implementations demand data verification before classifying an individual's thoughts or posts as suicidal or not. In order to perform better, a new model combining LSTM and NLTK has been developed.

Tadesse [24] was to showcase ongoing research on the automated identification of suicide messages. They use DL and ML on Reddit using social media to combat the early identification of suicidal thoughts. They use an LSTM-CNN combination model for this reason, evaluating and contrasting it with different classification methods. Their investigation shows that combining word embedding techniques with neural network architecture may yield the most relevant categorization results. Furthermore, their findings bolster the potency and capacity of deep learning architectures to construct a successful suicide risk assessment model across a range of text categorization tasks.

Arya, V. [25] When it comes to Facebook status sentiment analysis, Naive Bayesian has good accuracy in some situations. When hybrid classifiers are compared to traditional classifiers, they outperform traditional ones in terms of overall performance and accuracy. This is seen in the f-measures used in sentiment analysis. Framingham reported a 56% prediction accuracy for long-term cardiovascular disease. ML took into account novel topics such as drug misuse in the research experiment. In studies, almost 90% of medications are undetectable. According to Carnegie Mellon, automated drug discovery may cut costs by up to 70% practically.

Chadha [26] to help prevent the same by making use of social media as a resource. Suicidal ideation-related elements of Twitter, a social networking site, are used to gather data. The preprocessed tweets are transformed into probabilistic values based on the semantics of the detected characteristics, making them suitable for use using ensemble learning and ML methods. To predict and detect patterns of suicidal thoughts, several machine learning techniques were used to the data, including Decision Tree, Logistic Regression, Bernoulli NB, Support Vector Machine, Multinomial NB. To further determine improvements, the suggested work is assessed using ensemble methods such as AdaBoost, Random Forest, and Voting Ensemble.

Narynov [27] create a machine learning model, implement both teacher-and student-led teaching strategies, and then determine which algorithm is best suited for the task of classifying if a text contains references to suicide through comparative analysis. Their study aids in recognizing depressing information that might inspire suicide and assists those who may benefit from it in getting the assured assistance of in recognizing depressing information that might inspire suicide. They can conclude that K-means (Unsupervised) with tf-idf offers remarkable results, which are just 4% lower in f1-score and precision. This is because K-means Supervised obtained the highest result for 95% of f1-score for Random Forest (Supervised) using tf-idf model.

Ramachandran [28] demonstrates that machine learning algorithms such as logistic regression can accurately predict a Twitter user's actual suicidal inclination by using a variety of machine learning approaches and analyzing and evaluating their efficacy for suicidal tendency identification.

Vesangi [29] examine the several techniques and strategies that might be employed to determine a person's depressive inclination. In terms of suicidal instances, it also offers a thorough grasp of the several sources of data that might be obtained. It might be beneficial for the innovation to provide a multifaceted approach that can identify this inclination and notify friends, family, or close ones in advance. In addition, this study outlines a number of methods that may be used to identify suicidal tendencies more accurately, including speech recognition, text pattern identification, daily physical activity analysis, facial expressions, and more. Lastly, an outline for the future study is provided by extrapolating the restrictions and future scope.

Paul, P. C [30] provide a methodical examination Depressive symptoms on websites and apps data in their research. The development of NLP in the Bengali language made this possible. CNN+GRU performs the best among the applicable deep learning and machine learning on their dataset in terms of f1-score, accuracy, precision and recall. Together, CNN and GRU outperformed some earlier studies with an accuracy of 87.11%. They plan to examine transformer-based models in the future and expand the dataset to its maximum potential.

Pachouly [31] applied ML techniques to determine—from their tweets—who could be depressed on Twitter. For this goal, they trained and tested predictive algorithms to identify is sad or not using characteristics collected from his/her behaviors inside the tweets. On a scale of 0 to 100%, classification machine methods are used to train and categorize it in various phases of depression. Tweets were used as the data source, and machine learning classification techniques were used to determine whether or not the tweeter was depressed. Predictive technique for early diagnosis of mental diseases such as depression in this way. The primary contribution of this study is the investigation of a feature's neighborhood and its influence on determining depression severity.

Gupta [32] examined Reddit postings to find individuals with mental health issues who may be about to self-harm or take their own lives. Six distinct divisions approaches are used in the construction of predictive systems throughout this process, and sentiment analysis is used to extract characteristics to express the emotional thoughts of an internet user. The Naïve Bayes classifier demonstrated a positive signal in resolving the suicide risk assessment task, emerging as a reliable classifier with a 71.40% accuracy rating.

Renjith [33] created and tested a technique to examine the psychological well-being of diverse Reddit users. They created this model with the goal of lowering the suicide rate in society by identifying suicidal intents in people's social media interactions. They were able to examine submissions by assigning weight to the data areas that showed whether or not the submission was suicidal using the LSTM-attention-CNN combination model. The results of our suggested model were likewise shown to outperform the alternative basis model, with 92.6% F1-score and 90.3% an accuracy. Their next research will focus on trying out various processes of attention, particularly Hierarchical Attention Networks, to determine how document organization may impact the functionality of their suggested categorization model.

Aldhyani [34] presented an experimentally-based methodology for developing a a technique for detecting suicidal thoughts using Reddit data sources that are publicly available. Using textual and LIWC-22-based characteristics, two tests were conducted to categorize Using the XGBoost model with the CNN and Bidirection memory model, social posts are classified as suicidal or not. Exam results were compared, and it was found that the CNN–BiLSTM model performed better when textual features were used. It achieved 95% detection rate for suicidal ideas, compared to 91.5% accuracy for the XGBoost model. Conversely, XGBoost outperformed CNN–BiLSTM when LIWC features were used.

Swain [35] proposed a sentiment analysis approach utilizing supervised learning to detect suicidal thoughts in tweets. As a prompt identification and alarm system for suicidal inclinations, this can be advantageous to society. The suggested approach leverages machine learning models and a variety of Python language modules for opinion mining.

Basu [36] demonstrated how the application of (NLP) and Machine Lear (ML) methods may identify individuals. A dataset that represented the various emotions and circumstances included in the social media posts was employed by them. Here, they classified the postings as suicidal or not used of machine learning (ML) techniques, such as XG-Boost, Naive Bayes, Random Forest, SVM, Gradient Boost and Decision Tree. Interestingly, the SVM method proved to be the most successful, attaining a f1-scorc of 0.93 and an accuracy of 92.67%. The Gradient Boost algorithm, on the other hand, produced the worst results.

Heckler [37] suggested a taxonomy for classifying the machine learning methods investigated in this field as well as one for emphasizing the difficulties in the field right now. Suicidal ideation has garnered growing interest in recent times, according to this research. Studies have often used text analysis and social media data exploration to look at the language used by people who may be suicidal. Furthermore, it appears that deep learning models are now popular in this field. Subsequent research on suicidal ideation should look at proactive, general models that don't rely on users' self-reports.

Chandra [38] presented an overview of several methods, including deep learning and machine learning, that are utilized to automatically recognize thoughts of committing suicide in online social network data. This study also discusses the kinds of characteristics that are employed and the extraction of features techniques for suicidal thought detection. This paper discusses the flaws of the present studies, compares the various methods for detecting suicidal ideation, and offers recommendations for future research in this field.

Yao [39] was to retrieve data that was not included in the sample but related to the relationship between opiate usage and suicidal thoughts. The convolutional neural network was the top classifier, with an F1 score of 96.6%. At least 90% of all models' classification results were obtained for at least one set of input combinations. It is less beneficial to use NN classifiers or conventional techniques to predict suicidal attitudes forecasting data that is not in the sample for postings that contain both indicators of opioid addiction and suicidal thoughts since Conventional approaches yielded more false negatives and NN classifiers more false positives. They show that NNs can be used as a predictive instrument to a non-sample target using a model constructed tool to a target that is not a sample features, we want to be able to discern in the out-of-sample target.

2.3 Comparative Summary and Analysis

Nowadays, the majority of psychologists use academic interviews and questionnaires to identify suicidal tendencies [40]. The voluntary nature of the entire procedure is the fundamental obstacle, even if there are numerous additional problems in deducing the mental condition of a patient from their investigation response. Inadequate mental state is viewed as a personal concern that must be managed by the individual themselves, therefore taking the first step is things that might make a suicide or sad person may hesitant to do [41]. Therefore, unless the patient is ready to volunteer, even if there are numerous professionals accessible to aid mentally sick people, they cannot truly reach out to the patients. Table 1 presents an overview of earlier research.

Table 2.3.1. Assessment with earlier research

Num	Authors	Algorithm	Accuracy
1.	Drus et al. [19]	Opinion-lexicon method	----
2.	Birjali et al. [20]	IB1, J48, CART, SMO, Naive Bayes	SMO
3.	Ji, S. et al [21]	Current suicidal ideation detection (SID) methods	_____
4.	Abdulsalam et al [22]	CNN	86.6%
5.	Dandekar et al [23]	NLP	______
6.	Tadesse et al [24]	NB, RF, XGBOOST, SVM, LSTM-CNN	(93.8%) LSTM-CNN
7.	Arya, V. et al [25]	SVM; Naïve Bayesian classifier	Naïve Bayesian
8.	Chadha et al [26]	Multinomial NB, Decision Tree, Logistic Regression, Bernoulli NB, SVM	79.30% (SVM)
9.	Narynov et al. [27]	Random Forest, K-means	RF (95% of f1-score)
10.	Ramachandran et al. [28]	Logistic Regression, SVM, Random Forest, Multinomi al NB	Logistic Regression (76.3%)
11.	Vesangi et al. [29]	____	____
12.	Paul, P. C et al. [30]	Deep learning, machine learning	CNN+GRU (87.11%%)
13.	Pachouly et al. [31]	NB, Linear classifier, and SVM	linear classifier (87.5%)
14	Gupta et al. [32]	SVM, J48, Naïve Bayes	Naïve Bayes (71.40%)
15	Renjith et al. [33]	LSTM-attention-CNN	90.3%
16	Aldhyani et al. [34]	CNN–BiLSTM, XGBoost	XGBoost (95%)
17	Swain et al. [35]	Supervised learning	-----
18	Basu et al. [36]	Decision Tree, Naive Bayes, Random Forest, XG-Boost, Gradient Boost, SVM	SVM (92.67 %)
19	Heckle r et al. [37]	Machine learning techniques	____
20	Chandra et al. [38]	Feature extraction methods	
21	Yao et al. [39]	CNN	F_1 score of 96.6%

2.4 The scope of the problem

On this topic, here have view several investigation articles. Their study involved the utilization of several deep and ML techniques to illness. One of the main issues in developing any machine learning project is data acquisition. Although we had some difficulties gathering data for this article, not much was gathered.

2.5 Challenges

- Data collection
- Model selection
- Model train

CHAPTER 3

Research Methodology

This study aims to identify the cause of suicidal propensity by using a mathematical prediction technique. It has the potential to save the lives of people with mental illnesses. We use a number of machine learning techniques that are covered in this work to analyze the data set in order to achieve this goal. Approaches to classification can aid in the prediction of major diseases. For this reason, we use data mining approaches to achieve our goal. This section looks at the entire data analysis process using sequential classification methods and the use of selected algorithms.

3.1 Research Instrumentation

This research presented an image of our research methodology here. We gathered our info through an internet connection. We worked utilizing the operating system Windows. Google's Colab notebook tools were used. "Colab" is a quick summary of a Google investigation instrument that might create using Python. Colab is the ideal instrument for education, data processing, and artificial intelligence since it allows anyone to write and execute any Python scripts within a web browser. A Jupyter server called Colab requires no setup provides unrestricted access to computer resources such as GPUs.

3.2 Data Collection Procedure/Dataset Utilized

People with heart illness were discovered through the use of machine learning. We thus had to gather a dataset. Data from patients are frequently included in datasets. We were unable to collect the data ourselves because of scheduling constraints. Using Kaggle, the dataset was acquired. For this reason, we selected reliable, excellent datasets from Kaggle that had less null and missing values. As a result, precision will increase. Figure 3.2.1 displays the dataset collection.

3.3 Statistical Analysis

Our collection was downloaded through the web. It's challenging to find Kaggle data online. For this, we had to conduct extensive web research. Finally, we were able to use Kaggle to get the dataset we had requested. Here are two kinds of datasets such as sentimental analysis and suicidal tendency datasets. Sentimental datasets have 13 and suicidal has two attributes. VThe dataset may be understood using a variety of techniques. The dataset is made easier to grasp with the help of the statistical analysis. In Sentimental, we employed five algorithms and one algorithm for suicidal. This fig. 3.3.1 is shown in Sentimental analysis and fig 3.3.2 is shown in suicidal tendency.

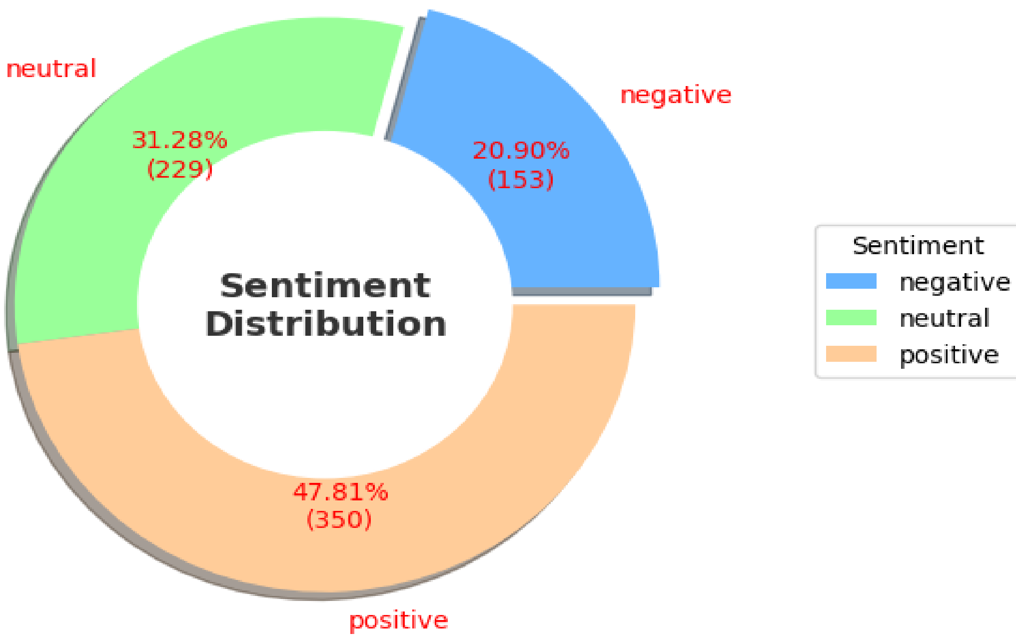

Fig.3.3.1 Sentimental dataset distribution

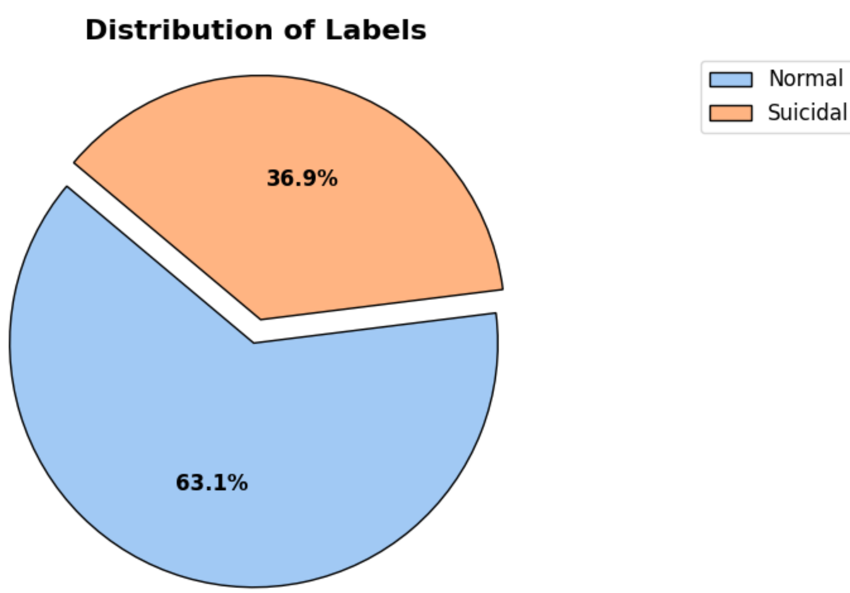

Fig.3.3.2 Suicidal dataset distribution

3.4 Proposed Methodology/Applied Mechanism

We aimed to identify it using a ML approach. To discover it, we tried a few different methods. Algorithms are Passive-Aggressive, random Forest, logistic regression, SVM, multinomial Naive Bayes, XGBClassifier. The procedure was done in many ways by us. Total unique values for Country: 33 which are USA: 188, UK: 143, Canada: 135, Australia: 75, India: 70, Brazil: 17, France: 16, Japan: 15, Germany: 14, Italy: 11, Spain: 6, South Africa: 6, Greece: 5, Czech Republic: 2, Netherlands: 4, Switzerland: 3, Portugal: 2, Austria: 2, Denmark: 2, Belgium: 2, Sweden: 2, Colombia, Scotland, Kenya, Jamaica, Ireland, China, Norway, Cambodia, Maldives, Peru, Jordan, Thailand : 1. Total unique values for Year: 14; They are distributed in 2023 (289), 2019 (73), 2020 (69), 2021 (63), 2022 (63), 2018 (56), 2017 (43), 2016 (38), 2015 (19), 2011(4), 2012 (4), 2013(4), 2014 (4), 2010 (3). Total unique values for Month are 12. The steps are illustrated in a responding flowchart form. Fig.3.4.1 Flowchart of data process. And table 3.4.1 shows about datasets.

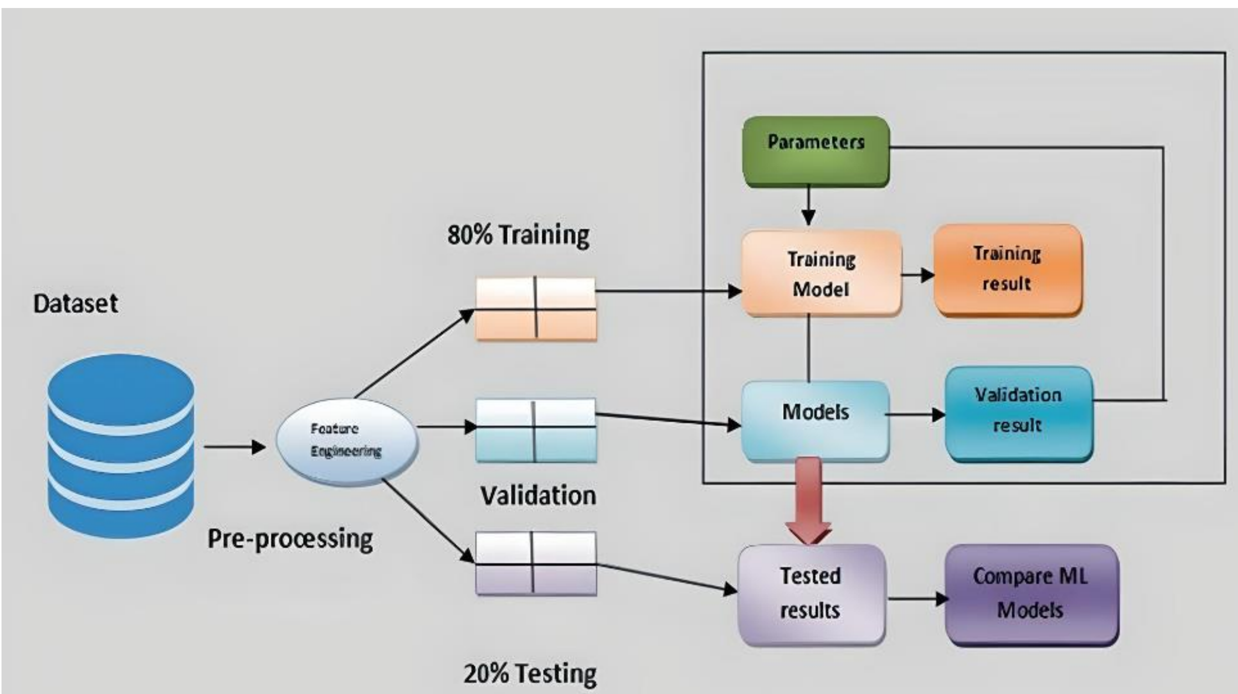

Fig. 3.4.1 Flowchart of data process

Here total dataset is divided two split, training datasets which is 70% and testing datasets which is 30%. Then The testing 30% datasets are divided into validation datasets about 10%.

Table 3.4.1 Categories of datasets

Categories	Sentimental dataset	Suicidal
Total data	732	1788
Attribute	13	2
Models	Passive-Aggressive, logistic regression, Random Forest, SVM, Multinomial NB	XGBClassifier

3.4.1 Data collection

The process of compiling information and data for a project is known as dataset collection. An algorithm is trained using a dataset, which is a collection of data in machine learning. The machine learning algorithm learns how to make predictions by using a dataset as an example.

3.4.2Data labeling

It is unsuitable to introduce a vast number of original details into a machine learning system and expect it to interpret it. Since these would result in immediate consequence, it is necessary pre-processing records involves a number of steps, one of which is data labelling. Throughout the data method, we recognize raw data and assign certain tags to it. Those tags specify the type of object to which the input corresponds, enabling learning in the model of ML and provide the most precise approximation.

3.4.3 Dataset preprocessing

Data preprocessing is getting the original information ready for machine intelligence techniques. It's the most important and initial step in constructing a model for algorithmic learning.

Not all of the data we see when working on a pattern recognition project is clean and well-prepared. Before data is used in any form, it also has to be cleaned up and classified. Thus, we employ the work of data preparation for all purposes.

Proper utilization of machine learning models is sometimes hampered by incompleteness and noise in actual data, as well as potential format issues. Preprocessing is a crucial step in order to clean up the information and get it ready for a categorization model, which improves the precision and effectiveness of ML algorithms.

It involves the subsequent actions:

- Obtaining the information
- Including collections
- Bringing in information
- Locating Absent Information
- Separating the dataset into testing and training sets using categorical data encoding
- Features scaling

We used certain preprocessing steps to our dataset. Our dataset was prepared in two phases. Two types of pre-processing methods were employed: regular scaling and null value management.

- It is known that values that are missing happen at random. Values that are missing can be managed by removing the columns or rows with zero values. A column may eliminate if more than half of the it contains rows that are null. Rows that have missing values in in one or more entries can also be eliminated.
- Additional scaling strategy is normalization, in which data are centered around the mean and have a standardized deviation. Consequently, the mean of the attribute is zero and the ensuing the variance of the dispersion is one.

Sentimental Analysis Datasets: There are some common words for positive signs. Fig 3.4.3.1 shows the picture of common words.

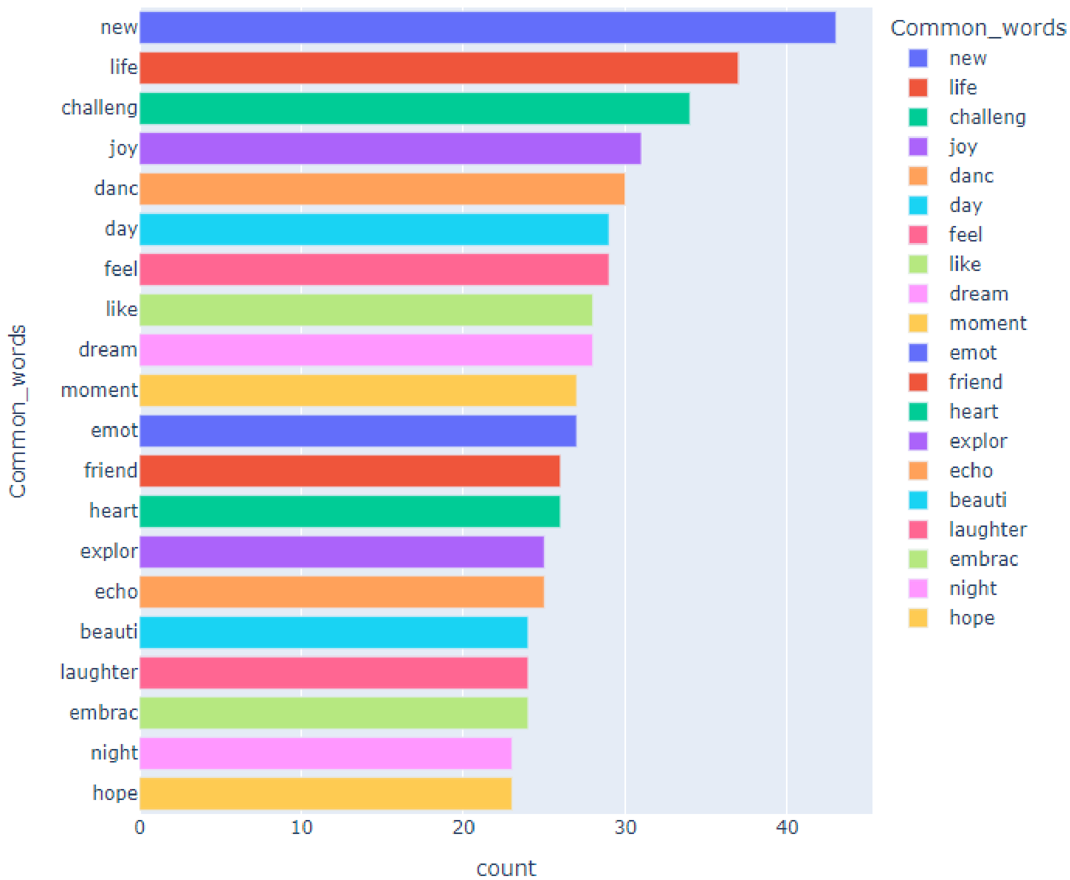

Fig. 3.4.3.1 Positive Common words for sentimental

Suicidal tendency Datasets: For suicidal dataset, we find out some words among normal and suicide people. Fig 3.4.3.2 shows the common words among them.

Fig 3.4.3.2 Difference between the common words normal and suicidal tweets

3.4.4 Performance measures

Our data is utilized to calculate the overall recall, accuracy, and F1 score—metrics prepared to evaluate the correctness and efficiency of architectural design. In this study, FP stands for false-positive and TP for true-positive. False-negative is denoted by FN and true-negative by TN. Our research indicates that the maximum outcome is around 99.51%.

Absolute Positive Precision through the process of dividing the total number of favorable results by the total quantity of outcomes, precision may be calculated. Recall is determined by multiplying the entire amount of real successes achieved by the projected positive class of results. A balance is struck between memory and accuracy via the harmonized mean, or F1 score.

The performance of the categorization task is evaluated using F1 scores, accuracy and recall.

- **Precision**

  One measure of a machine learning frameworks performance is precision, which evaluates the algorithm's favorable prediction's accuracy. Precision is computed the number of genuine positives divided by the overall amount of positive forecasts.

- **Recall**

  Recall is the sum of false negatives and true positives, determined using the number of affirmative observations in the data collection. An essential indicator of the model's capacity to identify the class labels is recall.

- **F1 scores**

A solitary assessment measure known as the F1 score aims to maximize recall while also making an effort to offset errors in accuracy. The arithmetic mean of accuracy is its definition and recall. F1 scores are used by data scientists to base on pre - defined predictive accuracy during model iteration stages by combining accuracy and recall into a single statistic. This allows teams to conduct hundreds of tests at the same time and objectively identify top-performing models. A simulation will achieve a strong F1 score if it has a high accuracy and recall. A model, however, will have a poor Highest accuracy if one element is low, and when the other is 100%.

- **Accuracy**

A measurement's accuracy is its distance from the real or acceptable value. The degree of accuracy between measurements of the same object is referred to as precision. Measurements with little error show excellent precision. When discussing accuracy in industrial instrumentation, one must consider the measurement tolerance of the apparatus, which determines the maximum and minimum amount of error that can occur under typical operating settings. One of the most crucial aspects of our measures is accuracy, which is crucial when utilizing data to inform choices.

$$\text{precision} = \frac{TP}{TP+FP} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

$$\text{recall} = \frac{TP}{TP+FN} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

$$\text{f1score} = 2\,X\ \frac{precision \times Recall}{precision + Recall} \quad \ldots\ldots\ldots \text{(iii)}$$

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

3.5 Implementation Requirements

The models listed were employed in our sentiment analysis: Passive-Aggressive, logistic regression, Random Forest, SVM, Multinomial NB. The highest Passive-Aggressive score of 72.78 %. In suicidal case, there is used in XGBClassifier model which accuracy is 96.08%.

3.5.1 Passive Aggressive Classifier

The Passive-Aggressive algorithms are a class of ML algorithms are unfamiliar to beginners and even intermediate machine learning aficionados. They are able to, however, be quite beneficial and effective in some situations. Online machine learning methods accept incoming data in sequential order and update the machine learning model part to part. This is extremely beneficial in circumstances where there is a large quantity of data and training the entire dataset is computationally infeasible due to its sheer size. We may simply state that an online-learning system will receive a training example, update the classifier, and discard the sample. A great example of this would be detecting bogus news on a social media platform like Twitter, where fresh data is generated every second. Passive-Aggressive algorithms are named thus because:

Passive: If the forecast is right, maintain the model without making any adjustments. In other words, the data in the case is insufficient to update the model.

Aggressive: If the forecast is wrong, make improvements to the model. i.e., a exchange the model might rectify it. Fig 3.5.1.1 shows the model of Passive Aggressive Classifier

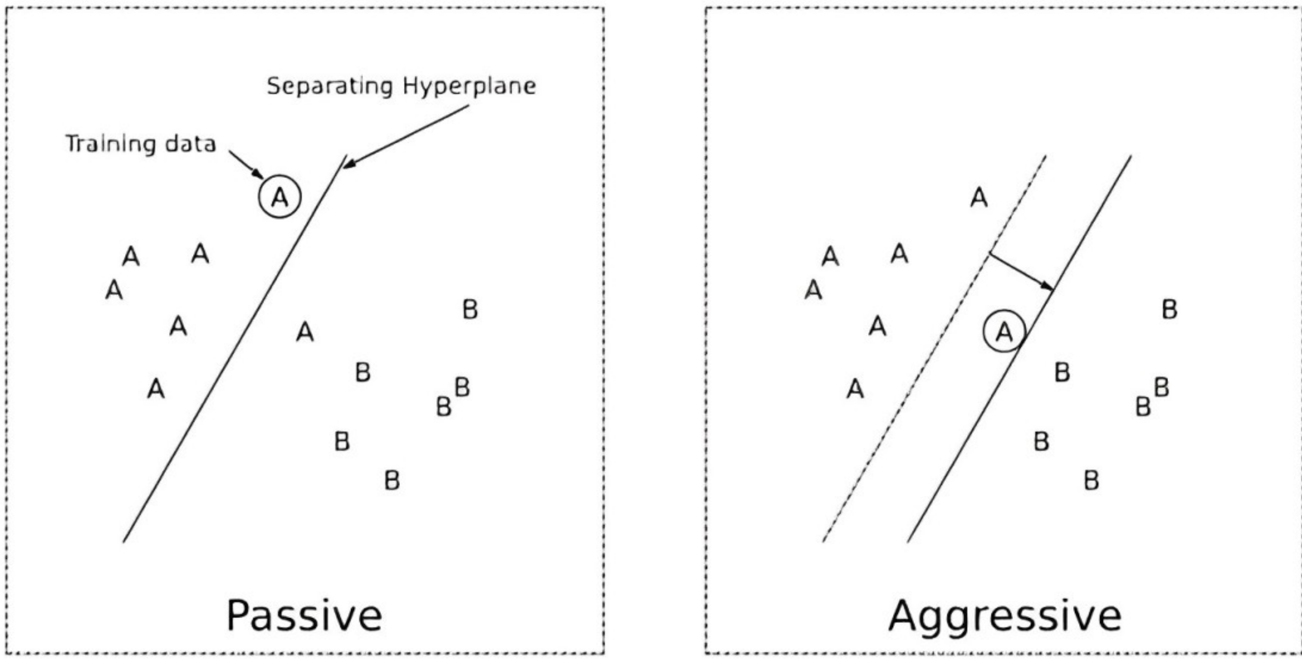

Fig. 3.5.1.1 Passive Aggressive Classifier

3.5.2 Logistic Classifier

When using accepting input as independent variables is the sigmoid function and outputs a probability value between 0 and 1, logistic regression is utilized for binary classification. An input falls into If the logistic function value above the 0.5 criterion, class 1 will apply; otherwise, it falls into Class 0. Its name comes from the fact that it is essentially a continuation of linear regression and is used mostly for classification problems. Predicting a categorical dependent variable's outcome is possible with logistic regression. Therefore, the outcome has to be a discrete or category statistic. It gives probabilistic values, which range from 0 to 1, rather than the exact values, which are 0 and 1. Logistic regression and linear regression are quite similar, save from how they are used. While classification problems can be resolved with logistic regression, linear algorithm is used to solve regression problems. Instead of fitting a regression line, a "S" shaped logistic function is used in logistic regression to predict two maximum values.

The logistic function curve displays the likelihood of several events. Logistic regression is a popular artificial intelligence methodology because it can generate probabilities and categorize new data using continuous and discrete datasets.

Observations may be classified utilizing a variety of data formats with the use of logistic regression, which also simplifies the process of determining the most effective criteria for the classification.

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

$$\frac{\mathbf{y}}{\mathbf{1-y}}\text{; 0 for y= 0, and infinity for y=1}$$

$$\boldsymbol{log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n}$$

Fig 3.5.2.1 shows of logistic regression classifier.

Three are present to form of logistic regression depending on the groups:

- Binomial: Using a binary regression model, the response variable can only have one of two possible kinds,
- Multinomial: In a multinomial regression analysis, the response variable might be any of three or more unordered types, for example, "sheep," "dogs," or "cats." A response variable with three or more ordered classes, such "low," "medium," and "high," when employing ordinal logistic regression.

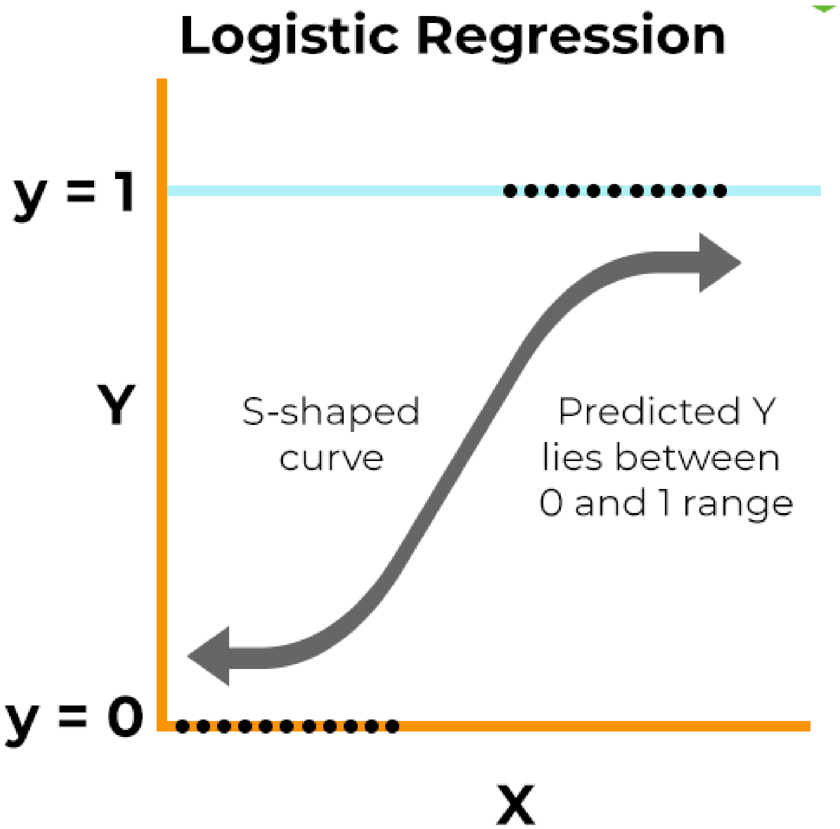

Fig 3.5.2.2 shows of RF.

3.5.3 RF Classifier

The popular ML algorithm Random Forest is used in the method to supervised learning. It might be used for resolve regression and classification difficulties in algorithmic learning. Its foundation is in the concept of supervised approaches, which integrate many categories to handle complicated problems and increase accuracy. As the name implies, "Random Forest" is a classification technique that selects the mean to increase the dataset's expected accuracy while offering a selection of decision trees for low-quality input. Rather of on solely on one decision tree, the rf aggregates the predictions from all of the decision trees and predicts the final result based on the simple majority of forecasts.

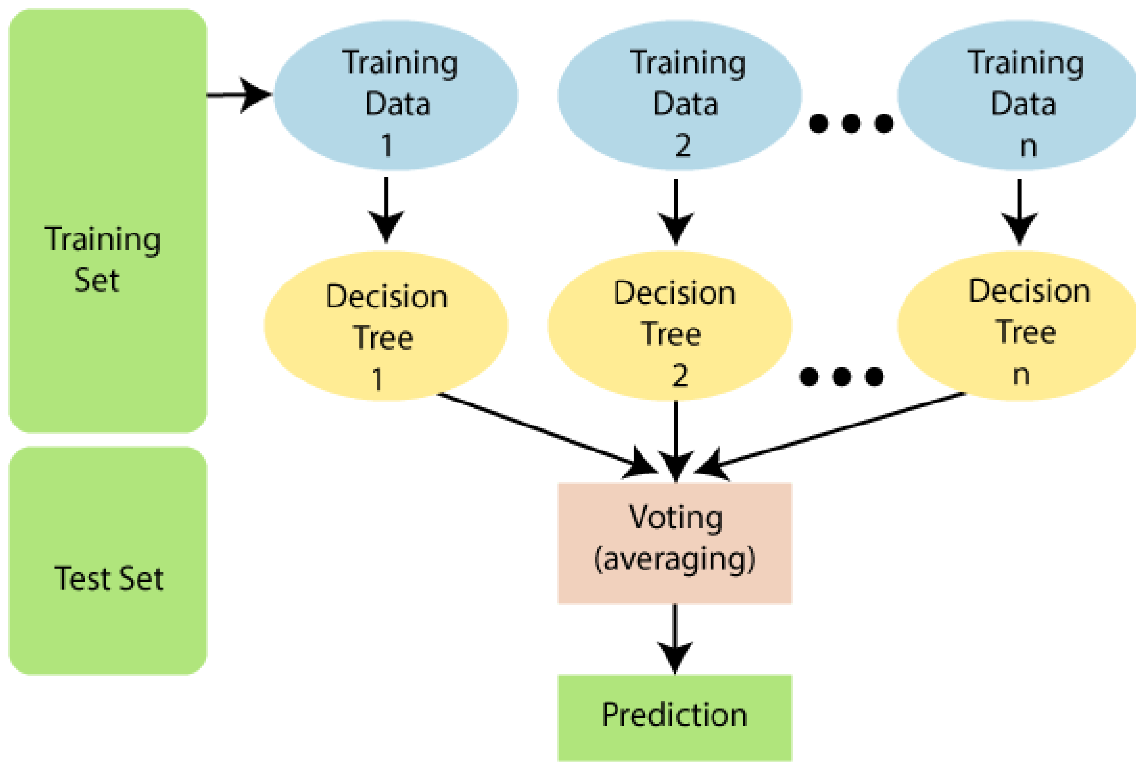

Fig 3.5.3.1 Random Forest

Certain decision tree techniques may yield the right answer while others may not because the random forest uses many trees to forecast the category of the dataset. Nevertheless, when considering the trees collectively, they accurately forecast the outcome. Thus, the following two conditions must be met for a more reliable Classification algorithm:

- In order for the algorithm to forecast precise results rather than guesses, the dataset reports and tests must include some real values.
- There should be very little correlation between the estimates for each tree.
- It anticipates output correctly and needs less time to train than other algorithms.
- It also performs well with large datasets.
- Lastly, accuracy can be maintained even in cases when a substantial portion of the data is missing.

3.5.4 SVM Classifier

Regression and classification are handled via a supervised machine learning technique known as SVM. Regression problems are best used, nonetheless, in classification problems. The main objective of the SVM method is to find the optimal hyperplane in an N-dimensional space so that data points may be divided into different feature space classes. The hyperplane makes an effort to keep the maximum possible buffer between the closest points of different classes. The number of features determines the hyperplane's dimension. Essentially, the SVM method looks for the best line or decision boundary that may split n-dimensional space into classes so that a new data point can be placed in the right category thereafter. A hyperplane is the name given to this ideal decision boundary. Using SVM, the vectors or extreme points are chosen to build a hyperplane. The methodology referred to as SVM is named after these extreme situations, which are called support vectors. Fig 3.5.2.2 shows of SVM. Examine the following graphic, which uses a hyperplane or decision border to divide two categories into different categories:

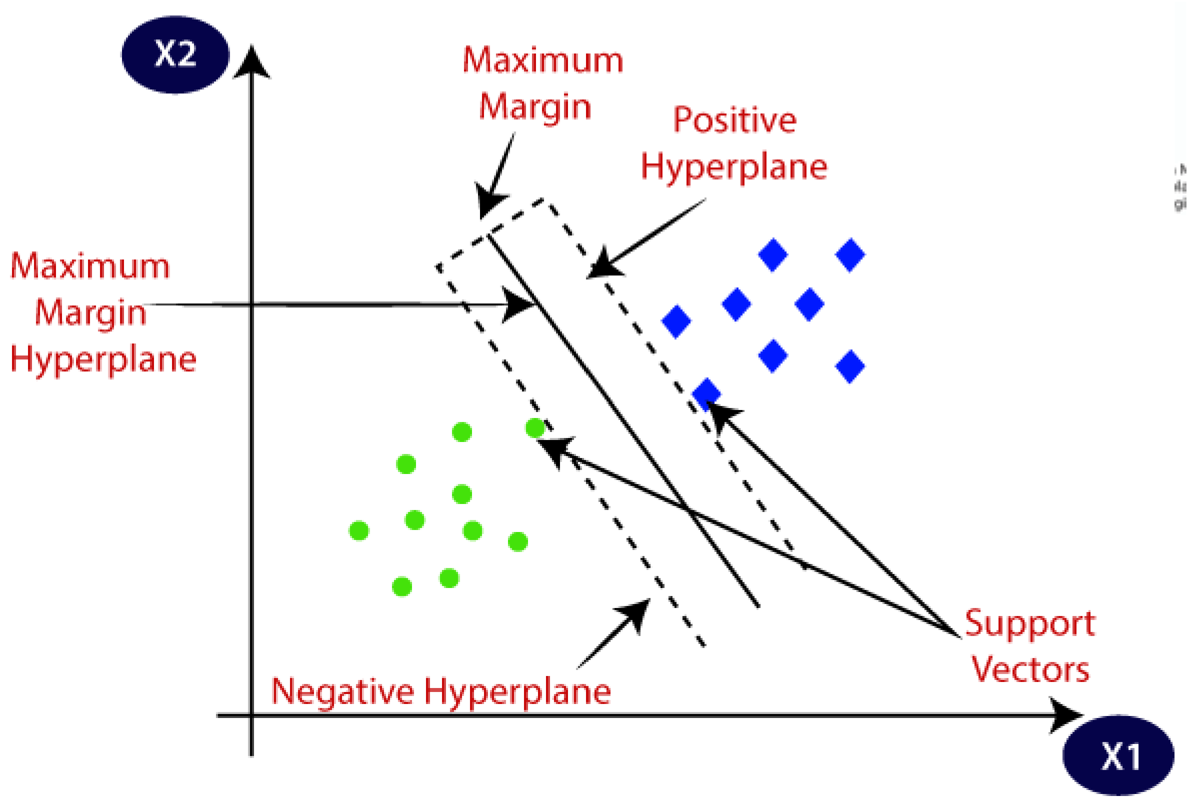

Fig 3.5.4.1 SVM

3.5.5 Multinomial NB

When using discrete features for classification, the multinomial Naive Bayes classifier works well. For the multinomial distribution to work, integer feature counts are often needed. In actuality, though, fractional counts like tf-idf could also function. Naive Bayes is a set of probabilistic algorithms depending on Bayes' Theorem. It's "naive" because it assumes feature independence, which indicates that the existence of one characteristic has no effect on the presence of others. A probabilistic classifier called Multinomial NB determines the likelihood distribution of the input text, making it ideal about information with distinct frequency or quantity of occurrences in NLP problems. Fig 3.5.2.2 shows of multinomial NB

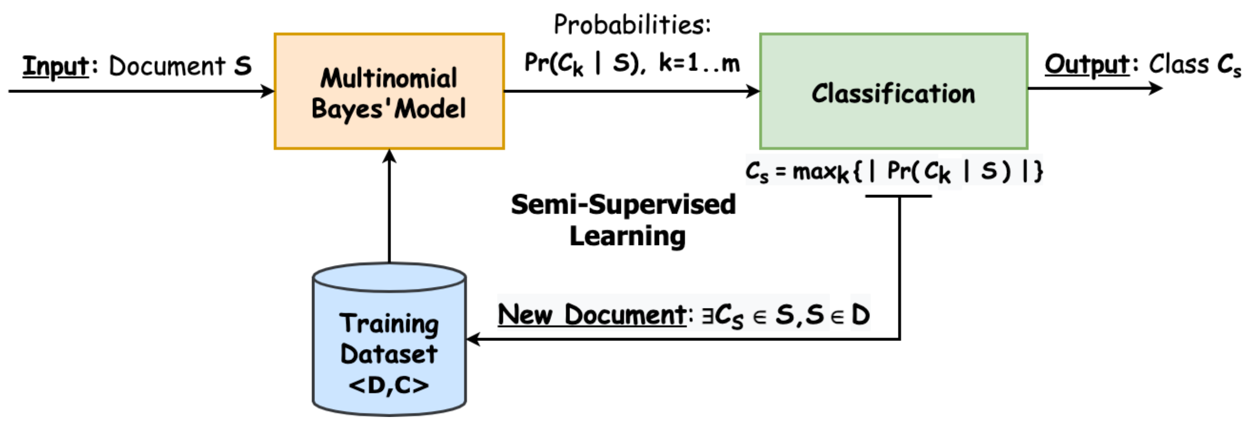

Fig 3.5.5.1 Multinomial NB

3.5.6 XGB Classifier

A more effective and scalable toolkit for distributed gradient boosting, XGBoost is used to train machine learning models. It is an ensemble learning technique that creates a more reliable prediction by combining the predictions of many weak models. Because of its remarkable success in a range of machine learning tasks, including regression and classification, and its capacity to handle big datasets, XGBoost, or "Extreme Gradient Boosting," has become the most widely used and accepted machine learning technique. Effectively managing missing values is one of XGBoost's unique features, allowing it to handle real-world data with missing values without requiring a lot of pre-processing. Moreover, XGBoost supports parallel processing, which enables you to train models on large datasets quickly. There are several uses for XGBoost, such as click-through rate prediction, recommendation systems, and Kaggle competitions. Furthermore, it is highly adaptable, enabling performance improvement through the fine-tuning of several model parameters. It stands for Extreme Gradient Boosting, a concept suggested by academics at the University of Washington. It is a C++ software designed to maximize gradient boosting training.
Fig 3.5.6.1 shows the diagram of XGBoost.

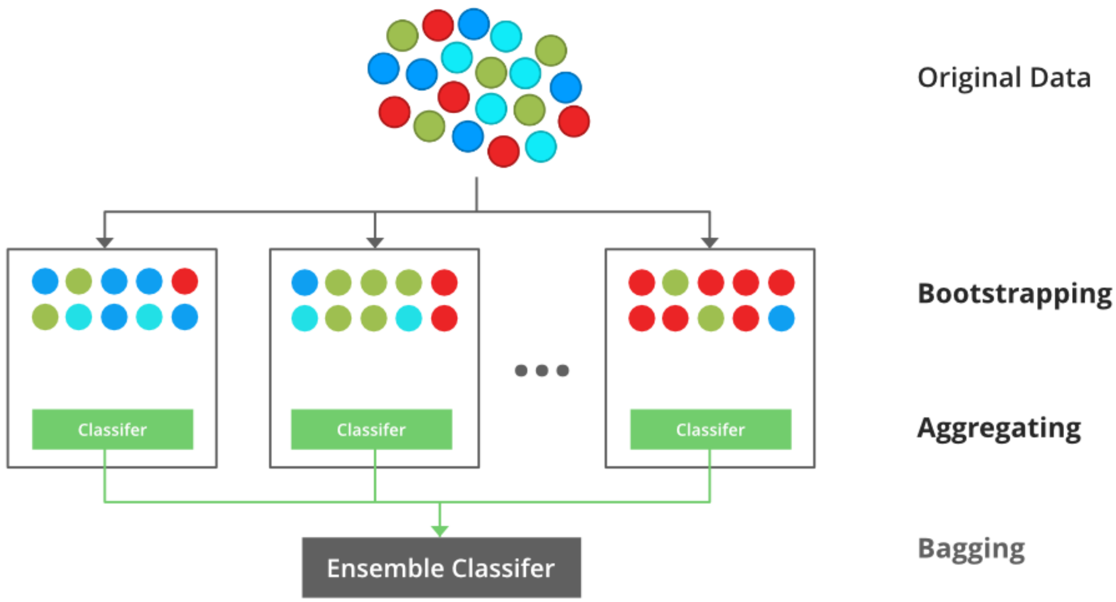

Fig 3.5.6.1 XGBoost process

CHAPTER 4

Experimental Results and Discussion

4.1 Experimental Setup

In this study, we analyze the effects of the last phase of the application employing actual data. We can attain a somewhat accurate performance by using our estimates. According to our examination, we have worked with approximately total 2520 pieces of data. Each data set we used, we gathered 13 attributes for sentimental datasets and 2 attributes for suicidal. The datasets are collected from Kaggle platform. Following that, Google Collab's machine learning tool was used to train the data.

4.2 Experimental Results & Analysis

Here, every model is currently carefully examined. First, we had to create the model and choose the model with the highest precision. Furthermore, the data was standardized to achieve accurate and consistent results. Table 4.2.1 shows the accuracy.

Table 4.2.1 Applied models accuracy

Sentimental analysis		
No	Algorithm Name	Accuracy Score
1	Passive-Aggressive, logistic regression, Random Forest, SVM, Multinomial NB	72.78%
2	Passive-Aggressive, logistic regression, Random Forest, SVM, Multinomial NB	63.26%
3	Passive-Aggressive, logistic regression, Random Forest, SVM, Multinomial NB	65.30%
4	Passive-Aggressive, logistic regression, Random Forest, SVM, Multinomial NB	59.86%
5	Passive-Aggressive, logistic regression, Random Forest, SVM, Multinomial NB	61.90%
Suicidal Tendency		
	Algorithm Name	Accuracy Score
6	XGBoost	97.08%

The capability of a model to identify every pertinent example inside a dataset. The recall is specified technically as the percentage of real positives split by the entire number of positive cases plus the total number of false negatives. Recall and precision refer to a classification model's capacity to recognize only those data items. Mathematically, the definition of precision is the quantity of false positives split by the total of the sum of the real positives and the false positives. Fig 4.2.1 shows the recall, precision, F1 score and fig 4.2.2 shows the confusion matrix for best Passive-Aggressive algorithm to sentimental tendency dataset.

	precision	recall	f1-score
negative	0.84	0.81	0.83
neutral	0.74	0.56	0.64
positive	0.68	0.83	0.75
accuracy			0.73
macro avg	0.75	0.74	0.74
weighted avg	0.73	0.73	0.72

Fig 4.2.1 The recall, precision, F1 score for Passive-Aggressive algorithm

N is the number of class labels in the confusion matrix, which is a N x N matrix utilized for evaluation the performance of a grouping technique. The ML model's predictions and the actual values of the targets are compared using a matrix.

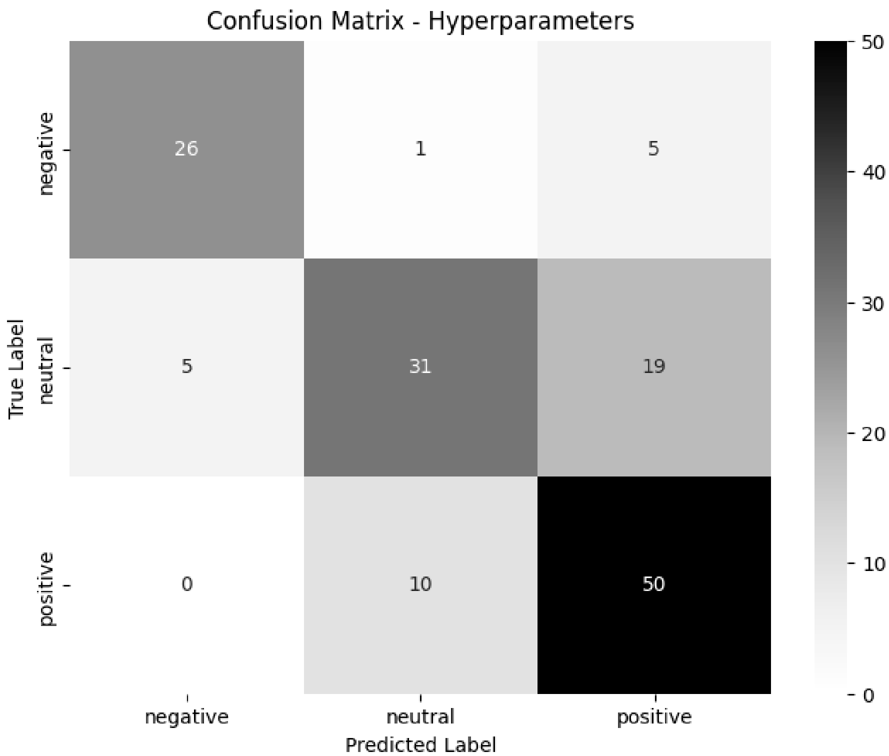

Fig. 4.2.2 Confusion matrix for Passive-Aggressive model

We divided the entire dataset in half to train the data. The training dataset is the first, while the test dataset is the second. We set aside 30% of the data for testing and 70% for training. We have used XGBoost algorithm for suicidal datasets which has suicide and normal types. Fig 4.2.3 shows the recall, precision, f1-score.

	precision	recall	f1-score
0	0.96	0.97	0.97
1	0.96	0.94	0.95
accuracy			0.96
macro avg	0.96	0.96	0.96
weighted avg	0.96	0.96	0.96

Fig 4.2.3 The precision, recall, f1-score of XGBoost

ROC curves are graphs that indicate classifier performance by graphing the true positive and unjustified positivity rates. The region covered by the ROC bend (AUC) is a performance metric for machine learning algorithms. ROC curves graphically represent the statisical accuracy of classifier selection; however, the graph's initial application was in signal detection. ROC curves, when used in conjunction with AUC, demonstrate how effectively an algorithm classifies objects based on AUC's invariance with the class under consideration. A ROC curve focuses on identifying the faults and advantages that classifiers use to structure classes, making ROC graphs helpful for comparing two classes in a diagnostic test that determines if a condition exists or not in an individual class. Fig 4.2.4 shows the ROC curve and fig 4.2.5 shows confusion matrix for XGBoost algorithm.

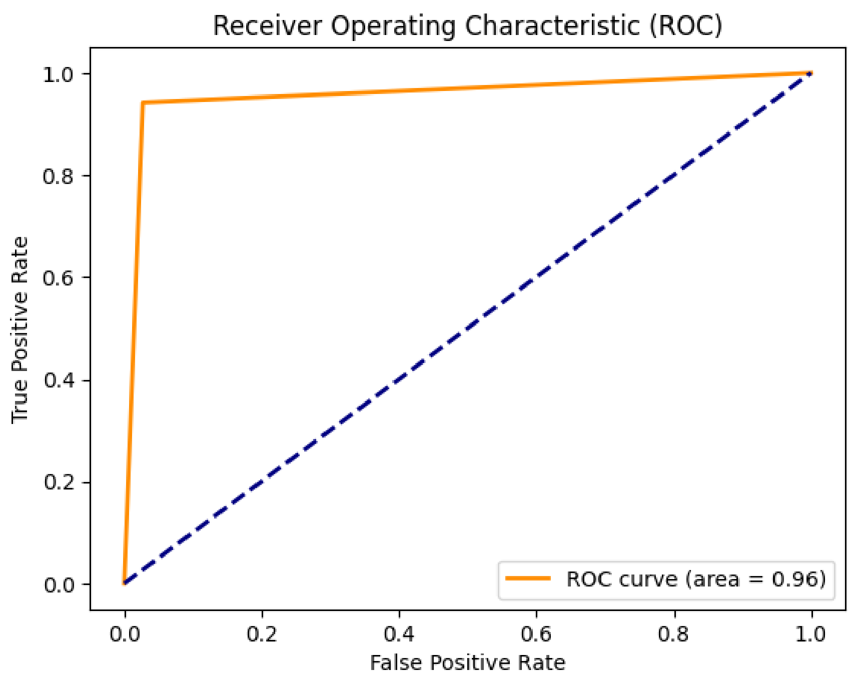

Fig 4.2.4 ROC curve for XGBoost model

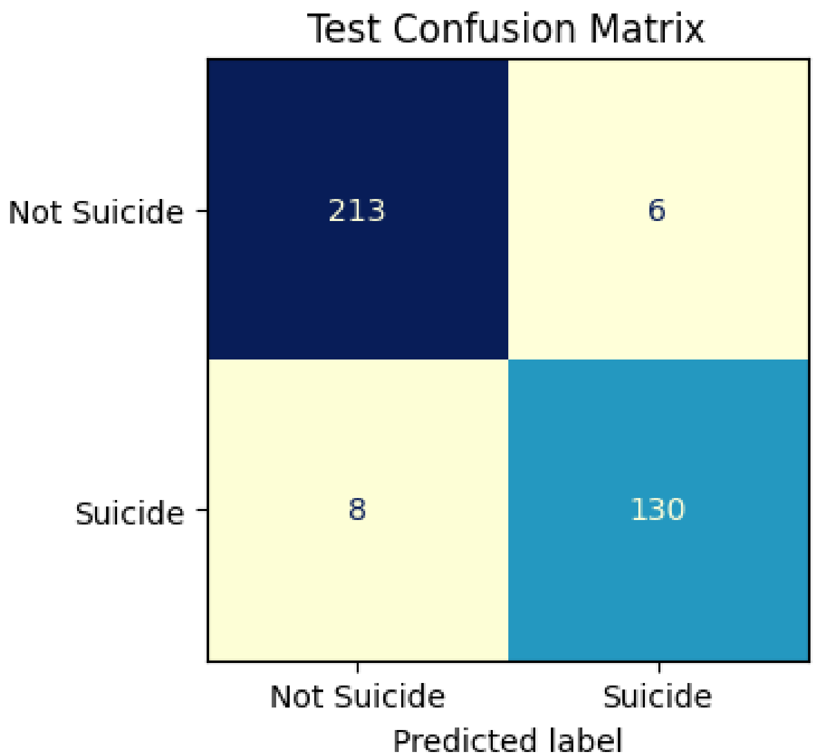

Fig 4.2.5 Confusion matrix for Passive-Aggressive model

4.3 Discussion

Researchers from several domains have focused on recognizing suicidal thoughts by analyzing linguistic and psychological signals, as well as other aspects. Social media posts offer insights into people's lives, emotions, and psychological states. Many people struggle to discuss their tales and emotions in person, leading them to create blogs about their sentiments or suicidal intentions. Regretfully, suicide posters are frequently disregarded or ignored. Large-scale suicidality screening can benefit from this knowledge.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

In the period of rapid technological advancement, a vast number of microblogging sites have emerged throughout the past ten years. Numerous social media networks are becoming more and more well-known every day. Social media platforms' primary factor in their popularity is their ability to encourage user interactions. Individuals are now openly and fearlessly voicing their feelings on social media. People are participating on social media, which generates a massive quantity of data every second. People discuss their feelings, their ideas about other people, their political viewpoints, and a variety of other topics on social media. Because these individuals are expressing their private thoughts and observations about their environment on social media platforms without realizing that they may go viral, it is possible to analyze these data sets of various people in order to understand their psychological behavior. Due to the diversity of their unique perspectives, different people have differing ideas about the same occurrence.

5.2 Impact on Society

In our study, we employed data about healthcare. It provided patient information. To decide how to treat patients, the system can assess data and use data mining techniques. Medical errors can be reduced, patient safety and outcomes may be enhanced, and unwanted practice variations can be eliminated when clinical decisions are supported by computer-based patient data. Strong modeling of data and evaluation technologies like data mining have contributed to the development of a knowledge-rich environment. This may improve the caliber of expert opinions. This automated scenario would replace current procedures when physicians choose treatments based on their gut reaction and previous experience. People will also be informed about the danger factors related to air pollution, workers in restaurants, water contamination, and so on. We can thus acquire an improved awareness of the environmental factors.

5.3 Ethical Aspects

While working on this project, there are a few ethical considerations that need to be made. so many people's personal information is in danger. Thus, it is our moral obligation to protect their privacy. Besides, we ought to remember:

- Not to put any patients in danger
- Security of private information
- Not to show genetic biases
- Designing studies objectively
- Accept liability for the study's results

5.4 Sustainability Plan

Almost perfect dependability was attained during our examination. In our nation, getting medical treatment is hampered by a number of factors. We have no access to money or medical professionals. Our automated technology enables prompt diagnosis and does away with the necessity for clinical testing. Thus, it would be extremely advantageous for individuals if we could make our project available to them.

CHAPTER 6

Summary, conclusion, recommendation and implication for future research

6.1 Summary of the Study

For a long time, depression has been treated as a single sickness with a set of diagnostic criteria, similar to other normal psychological state disorders. It typically co-occurs with anxiety or other psychological and physical illnesses, and it affects the way those who are afflicted feel and behave. According to a World Health Organization research, 322 million individuals globally, or 4.4% of the world's population —are estimated to suffer from anxiety. Social media communication is becoming increasingly important in today's environment. On social media platforms, they are open to sharing their ideas, narratives, mental health, needs, and other details. Recipients utilize email manuscripts and other types of social media comments to rectify errors and make the right decisions. People's texts are automatically processed on social media once they write digitally. Techniques used in the language communication process are used to infer people's mental behaviors. The highest accuracy from passive-Aggressive algorithm for sentimental analysis and XGBoost model gives 96% accuracy for suicidal tendency.

6.2 Conclusions

Among the recent issues in autonomous that arose with the emergence of social media platforms is sentiment analysis. Utilizing the wealth of information at our disposal, researchers and business leaders have looked for methods to automatically evaluate user attitudes and views posted on social networks. Suicide is defined as an intentional attempt on one's own life. An intricate combination of psychological, biological, social, spiritual factors and cultural leads to the multifarious incidence of suicide. Suicide is a symptom of fundamental distress induced by a variety of circumstances, including basic psychological disorders can produce psychological agony. Suicidal conduct comes in three forms: suicidal thoughts, suicidal planning, and suicidal attempts. There are finding 72% accuracy in Passive-Aggressive algorithms and 96% accuracy in XGBoost for suicidal cases. So, it can say that we find better accuracy for suicidal datasets.

6.3 Implication for Further Study

We will continue using more algorithms and datasets in an effort to improve performance and evaluate accuracy. We intend to carry out further research on this subject and create an Android app in the future to make it easier for individuals to detect their illnesses. We shall all profit from being able to predict one's own state.

Reference

[1] World Health Organization. (2014). Preventing suicide: A global imperative. Geneva, Switzerland. Retrieved from http://apps.who.int/iris/bitstream/10665/131 056/1/9789241564779_eng.pdf?ua=1&ua= 1

[2] https://vk.com/topic-78863260_30603285 - About VK Social Network

[3] Beck, A.T.; Kovacs, M.; Weissman, A. Hopelessness and suicidal behavior: An overview. JAMA 1975, 234, 1146–1149.

[4] Silver, M.A.; Bohnert, M.; Beck, A.T.; Marcus, D. Relation of depression of attempted suicide and seriousness of intent. Arch. Gen. Psychiatry 1971, 25, 573–576.

[5] Klonsky, E.D.; May, A.M. Differentiating suicide attempters from suicide ideators: A critical frontier for suicidology research. Suicide Life-Threat. Behav. 2014, 44, 1–5.

[6] Pompili, M.; Innamorati, M.; Di Vittorio, C.; Sher, L.; Girardi, P.; Amore, M. Sociodemographic and clinical differences between suicide ideators and attempters: A study of mood disordered patients 50 years and older. Suicide Life-Threat. Behav. 2014, 44, 34–45.

[7] World Health Organization. National Suicide Prevention Strategies: Progress, Examples and Indicators; World Health Organization: Geneva, Switzerland, 2018.

[8] DeJong, T.M.; Overholser, J.C.; Stockmeier, C.A. Apples to oranges?: A direct comparison between suicide attempters and suicide completers. J. Affect. Disord. 2010, 124, 90–97.

[9] De Choudhury, M.; Kiciman, E.; Dredze, M.; Coppersmith, G.; Kumar, M. Discovering shifts to suicidal ideation from mental health content in social media. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San José, CA, USA, 9–12 December 2016; ACM: New York, NY, USA, 2016; pp. 2098–2110.

[10] Catherine M McHugh, Amy Corderoy, Christopher James Ryan, Ian B Hickie, and Matthew Michael Large. 2019. Association between suicidal ideation and suicide: meta-analyses of odds ratios, sensitivity, specificity and positive predictive value. BJPsych open 5, 2 (2019)

[11] Deborah M Stone. 2021. Changes in Suicide Rates—United States, 2018–2019. MMWR. Morbidity and Mortality Weekly Report 70 (2021)

[12] James Vincent. 2017. Facebook is using AI to spot users with suicidal thoughts and send them help. The Verge (2017).

[13] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. NPJ digital medicine 3, 1 (2020), 1–11.

[14] Robert N Golden, Carla Weiland, and Fred Peterson. 2009. The truth about illness and disease. Infobase Publishing.

[15] Han-Chin Shing, Philip Resnik, and Douglas W Oard. 2020. A prioritization model for suicidality risk assessment. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 8124–8137.

[16] Radhakrishnan, R., Andrade, C.: Suicide: an Indian perspective. Indian J. Psychiatry 54(4), 304–319 (2012)

[17] Chestnov, O.: Public health action for the prevention of suicide: a framework. In: WHO Library Cataloguing-in-Publication Data, pp. 1–26. WHO Press, World Health Organization (2002)

[18] Luxton, D.D., June, J.D., Fairall, J.M.: Social media and suicide: a public health perspective. Am. J. Public Health Suppl. 2 102(S2), 195–200 (2012)

[19] Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, *161*, 707-714.

[20] Birjali, M., Beni-Hssane, A., & Erritali, M. (2017). Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. *Procedia Computer Science*, *113*, 65-72.

[21] Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. *IEEE Transactions on Computational Social Systems*, *8*(1), 214-226.

[22] Abdulsalam, A., & Alhothali, A. (2022). Suicidal ideation detection on social media: A review of machine learning methods. *arXiv preprint arXiv:2201.10515*.

[23] Dandekar, P. R., & Pethe, Y. S. (2021). Sentimental Analysis on Social Media by using Deep Learning. Annals of the Romanian Society for Cell Biology, 8254-8260.

[24] Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of suicide ideation in social media forums using deep learning. Algorithms, 13(1), 7.

[25] Arya, V., & Mishra, A. K. (2021). Machine learning approaches to mental stress detection: a review. Annals of Optimization Theory and Practice, 4(2), 55-67.

[26] Chadha, A., & Kaushik, B. (2021). A survey on prediction of suicidal ideation using machine and ensemble learning. The Computer Journal, 64(11), 1617-1632.

[27] Narynov, S. E. R. G. A. Z. Y., Mukhtarkhanuly, D. A. N. I. Y. A. R., Kerimov, I. L. M. U. R. A. T., & Omarov, B. A. T. Y. R. K. H. A. N. (2019). Comparative analysis of supervised and unsupervised learning algorithms for online user content suicidal ideation detection. J. Theor. Appl. Inf. Technol, 97(22), 3304-3317.

[28] Ramachandran, A., Gadwe, A., Poddar, D., Satavalekar, S., & Sahu, S. (2020, July). Performance evaluation of different machine learning techniques using Twitter data for identification of suicidal intent. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 223-227). IEEE.

[29] Baby, V., Vesangi, S., Regatte, S. R., Sudheshna, D., & Sarvani, D. Various Approaches for Suicidal Tendency Detection: A Literature Review.

[30] Paul, P. C., Ahmed, M. T., Hasan, M. R., Rajee, A., & Sultana, K. (2023). Analyzing depression on social media utilizing machine learning and deep learning methods. Indian Journal of Computer Science and Engineering, 14(5), 740-746.

[31] Pachouly, S. J., Raut, G., Bute, K., Tambe, R., & Bhavsar, S. (2021). Depression detection on social media network (Twitter) using sentiment analysis. Int. Res. J. Eng. Technol, 8, 1834-1839.

[32] Gupta, S., Das, D., Chatterjee, M., & Naskar, S. (2021). Machine learning-based social media analysis for suicide risk assessment. In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 2 (pp. 385-393). Springer Singapore.

[33] Renjith, S., Abraham, A., Jyothi, S. B., Chandran, L., & Thomson, J. (2022). An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. Journal of King Saud University-Computer and Information Sciences, 34(10), 9564-9575.

[34] Aldhyani, T. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. A. (2022). Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. International journal of environmental research and public health, 19(19), 12635.

[35] Swain, D., Khandelwal, A., Joshi, C., Gawas, A., Roy, P., & Zad, V. (2021). A Suicide Prediction System Based on Twitter Tweets Using Sentiment Analysis and Machine Learning. In Machine Learning and Information Processing: Proceedings of ICMLIP 2020 (pp. 45-58). Springer Singapore.

[36] Basu, S., Ahmed, M., Acharjee, P., & Saifuddin, M. (2024, May). Machine Learning-Driven Analysis of Suicidal Language on Social Media: A Dive into Natural Language Processing (NLP) Techniques. In 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) (pp. 1181-1186). IEEE.

[37] Heckler, W. F., de Carvalho, J. V., & Barbosa, J. L. V. (2022). Machine learning for suicidal ideation identification: A systematic literature review. Computers in Human Behavior, 128, 107095.

[38] Chandra, S., Bhattacharya, S., Banerjee, A., & Kundu, S. (2021). Suicide Ideation Detection in Online Social Networks: A Comparative Review. In Proceedings of International Conference on Innovations in Software Architecture and Computational Systems: ISACS 2021 (pp. 151-167). Springer Singapore.

[39] Yao, H., Rashidian, S., Dong, X., Duanmu, H., Rosenthal, R. N., & Wang, F. (2020). Detection of suicidality among opioid users on reddit: machine learning–based approach. Journal of medical internet research, 22(11), e15293.

[40] Ji, S., Yu, C. P., Fung, S. F., Pan, S., & Long, G. (2018). Supervised learning for suicidal ideation detection in online user content. Complexity, 2018.

[41] Vijaykumar, Lakshmi. "Suicide and its prevention: The urgent need in India." Indian journal of psychiatry 49.2 (2007): 81.

Final Rport (1) (1).pdf

ORIGINALITY REPORT

19%	16%	5%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to Daffodil International University Student Paper	6%
3	www.researchgate.net Internet Source	1%
4	Submitted to Visvesvaraya Technological University, Belagavi Student Paper	1%
5	Submitted to Liverpool John Moores University Student Paper	1%
6	Submitted to University of Surrey Student Paper	1%
7	Submitted to Alliance University Student Paper	<1%
8	Submitted to Damonte Ranch High School Student Paper	<1%