A Machine Learning Approach to Detecting Suicidal Tendencies in Adolescents

BY

Rafi Islam
ID: 202-15-14371
AND

Sultana Akter Sweety
ID: 202-15-14373

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

Md. Mahedi Hassan
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Md. Ferdouse Ahmed Foysal
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "A Machine Learning Approach to Detecting Suicidal Tendencies in Adolescents ", submitted by Rafi Islam and Sultana Akter Sweety 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 14/07/2024.

BOARD OF EXAMINERS

Dr. Md. Fokhray Hossain (MFH) **Professor** Department of CSE Faculty of Science & Information Technology Daffodil International University	**Board Chairman**
Amatul Bushra Akhi (ABA) **Assistant Professor** Department of CSE Faculty of Science & Information Technology Daffodil International University	**Internal Examiner 1**
Mr. Amir Sohel (ARS) **Sr. Lecturer** Department of CSE Faculty of Science & Information Technology Daffodil International University	**Internal Examiner 2**
14/7/24 **Dr. Shamim H Ripon (DSR)** **Professor** Department of the Computer Science & Engineering East West University	**External Examiner**

©Daffodil International University ii

DECLARATION

We hereby declare that this project has been done by us under the supervision of **MD. Mahedi Hassan, Lecturer, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree.

Supervised by:

mahedi
13.07.24

Md. Mahedi Hassan
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Md. Ferdouse Ahmed Foysal
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Rafi Islam

Rafi Islam
ID: 202-15-14371
Department of CSE
Daffodil International University

Sultana Akter Sweety

Sultana Akter Sweety
ID: 202-15-14373
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We are grateful and wish our profound indebtedness to **Md. Mahedi Hassan**, **Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Machine learning, Image processing, Natural Language Processing” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

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

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

ABSTRACT

Nowadays suicide has become a serious crime. People commit suicide for many reasons. Due to the advancement of social media, people post various types of posts before committing suicide. Understanding the environmental risk factors that affect suicide thoughts and behavior throughout time will be greatly aided by this study. We will collect various types of data from various online platforms and identify them with the help of machine learning models. We use some algorithms to find out the best accuracy. To find out the best accurate results we implement classifier such as Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Decision Tree Classification, Random Forest and Gradient Boosting Classifier. The models were evaluated based on their accuracy, precision, recall and F-1 score. Naive Bayes had the lowest accuracy at 73%, while Support Vector Machine (SVM) secured the top positions with 92.32% accuracy. This research highlights the importance of integrating advanced machine learning techniques into mental health care to facilitate early intervention and support for at-risk adolescents. By leveraging technology, we can enhance the effectiveness of suicide prevention strategies and ultimately save lives.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Overview

Suicide has become one of the most serious social health problems in modern society. According to a WHO report more than 700000 people are estimated worldwide committed suicide in 2023. The tendency to commit suicide is especially high among young people. They don't decide to act as deadly as suicide on impulse for no common reason. People can commit suicide for many reasons. Some commit suicide out of pride, some do it in vain, some do it to avoid responsibility, some do it out of shame. Also people commit suicide due to physical problems, various diseases. Every year many people around the world choose to commit suicide without any reason. According to previous studies, suicide is observed as the second highest cause of death among the young population with suicide rate is 10.5 per 100000 people. Most of the suicides occur in low-income countries. Based on these kinds of self-report measurements of indicator factors or variables, the majority of suicide research is conducted. Personal social data can help to find suicide detection and act as the key factor of suicidal tendencies. Nowadays people are publishing their daily activities on social media and its data is playing an important role in measuring their mental condition [1].

Social media brings people from all over the world together. People are connected to each other virtually through social media. People can now easily connect with each other through social media. This social media is currently the new way or means of expressing people's emotions, feelings, love and suicidal tendencies. Due to the massive advancement of mobile internet, people are sharing about their suicidal intention through social platform. These online posts help a lot in suicide detection. Many people, especially young people, post on social media about how to commit suicide. From these social media posts, it is known how many people commit suicide due to human pressure and immediate preventive measures can be taken in this regard. Nowadays social media serves as one of the main tools for suicide detection and studying them we prevent suicide [2]. Suicide detection is the first step in suicide prevention. Therefore, public awareness activities should be conducted to prevent suicide.

According to some research we see that, with the help of artificial intelligence and machine learning algorithms, we can easily identify suicides by searching user data. Which can help us to understand in our personal emotional state and establish an early intervention [3]. In this paper we will try to understand and identify suicide posts on different social media and detect suicidal tendencies. Here, we will apply machine learning techniques to create an algorithm that can detect suicidal ideas or thoughts based on data from social media site like Reddit, Twitter, Facebook. Our primary goal is to determine the risk of suicidal intent or behavior before suicide occurs, and to identify post-suicide risk factors.

1.2 Background and Present State

Adolescent suicide represents a critical public health issue globally, with rates showing concerning trends in recent years. In many parts of the world, suicide is one of the leading causes of death among young people aged 15-24 years. Traditional methods for identifying suicidal tendencies predominantly rely on subjective assessments, self-reporting, and clinical observations. While these methods provide valuable insights, they are often limited by their reliance on individuals' willingness to disclose sensitive information and the variability in clinicians' interpretations. Moreover, the dynamic nature of adolescent behavior and the stigma associated with mental health issues can further complicate accurate risk assessment and timely intervention. The emergence of machine learning presents a promising avenue for improving the accuracy and efficiency of suicide risk detection [4]. Machine learning algorithms can analyze vast amounts of data from diverse sources, including social media posts, online forums, and electronic health records, to identify subtle patterns and predictive markers of suicidal behavior. Previous research has demonstrated the potential of machine learning in enhancing risk prediction models by integrating multiple data points and detecting patterns that may not be discernible through traditional methods alone. Despite these advancements, challenges such as data privacy concerns, algorithm interpretability, and the ethical implications of automated decision-making remain significant considerations in the application of machine learning to mental health contexts [5].

1.3 Problem Statement

Adolescent suicide represents a critical public health issue globally, with rates showing concerning trends in recent years. In many parts of the world, one of the main reasons of death for young people between the ages of 15 and 24 is suicide. Traditional methods for identifying suicidal tendencies predominantly rely on subjective assessments, self-reporting, and clinical observations. While these methods provide valuable insights, they are often limited by their reliance on individuals' willingness to disclose sensitive information and the variability in clinicians' interpretations. Moreover, the dynamic nature of adolescent behavior and the stigma associated with mental health issues can further complicate accurate risk assessment and timely intervention. The emergence of machine learning presents a promising avenue for improving the accuracy and efficiency of suicide risk detection. Machine learning algorithms can analyze vast amounts of data from diverse sources, including social media posts, online forums, and electronic health records, to identify subtle patterns and predictive markers of suicidal behavior. Previous research has demonstrated the potential of machine learning in enhancing risk prediction models by integrating multiple data points and detecting patterns that may not be discernible through traditional methods alone. Despite these advancements, challenges such as data privacy concerns, algorithm interpretability, and the ethical implications of automated decision-making remain significant considerations in the application of machine learning to mental health contexts.

1.4 Objectives

This study aims to achieve several key objectives in the field of detecting suicidal tendencies among adolescents using machine learning approaches.

1. To Develop Machine Learning Models: Develop and optimize machine learning models capable of accurately detecting suicidal tendencies in adolescents. This involves exploring various algorithms and techniques to identify the most effective approach.

2. To Evaluate Performance: Evaluate the performance of the developed models using appropriate metrics such as accuracy, precision, recall, and F1 score. Compare the performance of different machine learning algorithms, including Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Decision Tree Classification, Random Forest, and Gradient Boosting Classifier.

3. To Identify Predictive Features: Identify key predictive features or markers of suicidal behavior from data sources such as social media posts and clinical records. This involves feature selection techniques and understanding which variables contribute most significantly to the models' predictive accuracy.

4. To Enhance Early Detection: Enhance early detection capabilities by integrating machine learning models into existing mental health assessment frameworks. This objective aims to provide healthcare professionals with timely and accurate insights into adolescent suicide risk, facilitating early intervention strategies.

5. To Contribute to Mental Health Research: Contribute to the advancement of mental health research by validating the effectiveness of machine learning in detecting suicidal tendencies. This includes discussing the implications of your findings for future research and clinical practice.

1.5 Scope and Limitations

We focuses on adolescents aged who may be at risk of developing suicidal tendencies. The primary goal is on implementing, developing and evaluating a machine learning model specifically designed for the early detection of suicidal tendencies in adolescents. Our focus is on leveraging anonymized mental health records, surveys, and relevant datasets to train and evaluate the effectiveness of the machine learning model. The scope encompasses the identification of key risk factors and warning signs associated with suicidal tendencies in adolescents, with a particular emphasis on cultural sensitivity and diversity considerations. We will explore various machine learning algorithms and techniques to develop a robust and accurate predictive model capable of identifying at-risk individuals. The deployment of the model will be within the context of real-time intervention systems, aiming to facilitate timely support and intervention for adolescents in need. Ethical considerations regarding privacy preservation, data security, and responsible use of sensitive information are integral to our project scope. Additionally, while our research primarily focuses on the development of the machine learning model, we recognize the importance of collaboration with mental health professionals and stakeholders to ensure the practical applicability and effectiveness of the model in real-world settings.

Limitations

Several limitations must be acknowledged in this study. While this study aims to contribute valuable insights into suicide risk assessment using machine learning, it is important to acknowledge certain limitations. :

- **Data Availability:** Dependency on the availability and quality of data from selected sources, which may affect the generalizability of findings.

- **Algorithmic Constraints:** Challenges in interpreting complex machine learning algorithms and ensuring the robustness of models across different demographic groups and contexts.

- **Ethical Constraints:** Potential ethical implications related to the use of personal data and automated decision-making in sensitive healthcare domains.

- **Scope of Clinical Application:** Practical considerations regarding the integration of machine learning models into existing clinical frameworks and the acceptance of automated risk assessments by healthcare professionals.

These limitations underscore the need for cautious interpretation of findings and highlight areas for future research to address these challenges effectively.

1.6 Report Organization

The suggested report has a comprehensive framework that takes readers through the methodology, decisions, and significance of the study. Every chapter has an individual purpose and helps to the document's overall structure and depth.

Chapter 1: Introduction

Provides an overview of the study, including the motivation, rationale, research questions, expected outcomes, and project management details.

Chapter 2: Literature Review

Reviews existing literature and research related to adolescent suicide, machine learning applications in mental health, and current methods for detecting suicidal tendencies. This chapter establishes the theoretical framework and contextualizes the study within the broader field of mental health research.

Chapter 3: Methodology

Describes the research design, data collection methods, and analytical techniques employed in the study. It outlines the process of developing and evaluating machine learning models for detecting suicidal tendencies among adolescents, including details on algorithm selection, feature extraction, and model evaluation criteria.

Chapter 4: Implementation

Implement the model different machine learning algorithm. How this model work and implementation of their work.

Chapter 5: Result and Analysis

Presents the performance of each machine learning model, key findings, and a comparative analysis.

Chapter 6: Impact on Society, Environment and Sustainability

The societal impact chapter explores what kind of effect will it have on the people of the society. It also discusses how people can get rid of it.

Chapter 7: Conclusion and Future Work

This final chapter serves as a synthesis of the entire report. It Summarizes the study, highlights key contributions, and suggests directions for future work.

This report layout ensures a systematic and thorough presentation of the research, providing a clear and logical progression from the introduction of the topic to the detailed analysis of results and their implications.

1.7 Summary

Adolescent suicide represents a critical public health concern, necessitating more effective methods for early detection and intervention. Current approaches predominantly rely on subjective assessments and clinical observations, which can be inconsistent and limited in their predictive accuracy. The emergence of machine learning offers promising opportunities to enhance suicide risk assessment by leveraging data-driven approaches and predictive modeling. This study aims to develop and evaluate machine learning models capable of identifying indicators of suicidal behavior among adolescents using diverse data sources. By comparing the performance of various algorithms and analyzing key predictive features, the research seeks to contribute to the advancement of suicide prevention strategies. While acknowledging the scope and limitations of the study, including data availability and ethical considerations, this research aims to provide valuable insights into the application of machine learning in mental health care. The subsequent chapters will delve deeper into the literature review, methodology, findings, and implications of using machine learning for detecting suicidal tendencies among adolescents.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

The literature review section provides a comprehensive overview of existing research and studies related to detecting suicidal tendencies, particularly focusing on adolescents, using machine learning approaches. This section provides an overview of key themes, methodologies, and findings discussed in the literature, setting the stage for a comprehensive analysis. It explores a wide range of scholarly articles, journals, and conference papers to understand the current landscape, methodologies, and findings in this field. Key topics covered include the prevalence of adolescent suicide, risk factors, warning signs, and existing approaches to suicide prevention and intervention. The review critically analyzes the strengths and limitations of previous studies, identifying gaps and areas for further research. Additionally, it examines the various machine learning algorithms and techniques employed in previous studies for detecting suicidal behavior, assessing their effectiveness and applicability. Cultural and demographic factors influencing suicidal tendencies are also explored to understand the complexities of this phenomenon across different populations. The literature review serves as the foundation for our research, informing the development of our methodology and guiding our approach to addressing the identified gaps in the existing literature. By summarizing the scope and relevance of existing studies, this overview sets the stage for a detailed exploration of related works and comparative analyses in subsequent sections of the chapter.

2.2 Related Work

Previous studies have explored various approaches to identify indicators of suicidal behavior through analysis of social media content and clinical data. Researchers have utilized machine learning algorithms such as Naive Bayes, SVM, Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting to analyze textual and behavioral data from adolescents. These studies have contributed valuable insights into the predictive power of different models and the effectiveness of integrating social media insights with traditional clinical assessments.

For instance,

Shaoxiong Ji et al [2018] the authors proposed a novel approach for detecting suicidal tendencies using a machine learning model and deep learning model. The authors used a dataset of 3549 suicidal ideation samples and compared the performance of various models. Their results indicate that Random forest model is better than with other model with an accuracy of 96.38% [6].

Soumyabrata Saha et al [2023] proposed a novel approach for detecting suicidal tendencies using a machine learning model. The authors collect data many online platform. The authors used 6 machine learning algorithm to detect suicidal tendencies. Their results indicate that SVM model is best with other models with an accuracy of 88.6% [7].

Marcel Adam Just et al [2017] the authors suggested a unique strategy for detecting suicidal tendencies using a machine learning model of neural representations. The authors used Gaussian Naive Bayes algorithm with high accuracy of 91% [8].

Gourab Karak et al [2023] employed a machine learning model to analyze textual data from therapy sessions, achieving promising results in identifying individuals at risk of suicidal thoughts. The authors used a dataset of 210 suicidal ideation samples and 97 attributes. The authors used 3 machine learning algorithm to detect suicidal tendencies. Their results indicate that the suggested Confusion Matrix model exceeded other models with an accuracy of 86.11% [9].

Asma Abdulsalam and Areej Alhothali [2022] the authors suggested a unique approach for detecting suicidal tendencies using a machine learning and deep learning model. The authors collect dataset many online platform compared the performance of various models. Their results indicate that SVM model surpassed other models with a 93.8% accuracy rate [10].

Hugo D. Calderon-Vilca et al [2022] the authors preferred an approach for detecting suicidal tendencies using a machine learning model. The authors used a dataset of 10000 suicidal ideation and compared the performance of various models. They also collect 880 real attempt of suicidal tendencies. Their results indicate that Decision Tree model is the best with other model with an accuracy of 98.4% [11].

Xiaolei Huang et al [2014] the authors expected a novel approach for detecting suicidal tendencies using a machine learning model and psychological knowledge. The authors use both real time and online platform data. The authors used a dataset of over 90000 suicidal and non suicidal ideation and compared the performance of various models. With an accuracy of 96.38%, their results show that the Random Forest model fared better than the other models [12].

Erik Cambria et al [2021] the authors suggested a unique strategy for detecting suicidal tendencies using a deep learning model. The authors collect dataset different platform like UMD dataset, Reddit SWMH, Twitter dataset. In UMD dataset, their results indicate that BiLSTM model is better with other models with an accuracy of 0.5694%. In Reddit SWMH dataset, their results indicate that RN model much better other models with an accuracy of 0.6474%. In Twitter dataset, their results indicate that RN model is the best with other models and this model accuracy is 83.85% [13].

Theyazn H. H. Aldhyan et al [2022] the suggested a unique strategy for detecting suicidal tendencies using a deep learning and machine learning techniques. The authors used a dataset of 232074 suicidal ideation samples and compared the performance of various models. Their findings show that the CNN-BiLSTM model performed better than the other models, with a 95% accuracy rate [14].

Moumita Chatterjee et al [2022] the authors suggested a unique strategy for detecting suicidal tendencies using a machine learning model and natural languages processing techniques. The authors used a dataset of 188,704 suicidal and non suicidal ideation samples and compared the performance of various models. Their results indicate that

Logistic Regression model performed much better than other models, with an accuracy rate of 87% [15].

Pratyaksh Jain et al [2022] the authors preferred an unique approach for detecting suicidal tendencies using a machine learning model and natural languages processing techniques. The authors used a dataset of 60000 suicidal ideation samples and compared the performance of various models. Their results indicate that Logistic Regression model is the best suitable with other models with 77.29% accuracy rate [16].

Jingfang Liu et al [2022] analyzed machine learning algorithm to detecting suicidal tendencies in adolescents. A dataset of 40,222 posts has been assembled by the researchers. The model achieved outstanding recognition accuracy of 80.61% for Logistic Regression model [17].

Mahmud et al [2023] the authors analyzed a machine learning algorithm to dectcting suicidal tendencies. The authors collect huge number of data from university student in Bangladesh. The authors distributed an online survey to collect data from online platform. The authors used 6 machine learning algorithm (Logistic regression, SVM, Random forest, K-nearest neighbors algorithm, Naive Bayes, Classification tree) to detect suicidal detection. With an accuracy rate of 79%, the suggested model performs as a thorough SVM [18].

Lie et al [2020] the authors proposed BiLSTM method to detect suicidal tendencies. The authours collect data many social media platform and train, test of them. The model achieve 78% accuracy [19].

Mobin et al [2024] the authors arrange a machine learning algorithm to detect suicide detection. The authors both use supervised data and unsupervised data. The authorrs use various maching learning alogorithm. After the implementation of all algorithm their result indicate Support Vector Machine (SVM) gave the best suitable accuracy 92% [20] .

Munner el al [2020] the authors use a machine learning algorithm to detection suicideal tendencies an adolescents. The authors used a 37373 dataset of twitter. tAfter apply of all algoritm their result indicate that logistic regressing is the best model with an accuracy rate of 91% [21].

Arjun et al [2023] the authors suggest a unique strategy of machine learning algorithm. The authors collect data form kaggle. The authors used 2000 sample of social media data and compare them various machine learning model. After the implementation of all machine learning algorithm their result indicate that GradientBoostingClassifier is the best of with other model with an accuray of 92.15% [22].

Chowdhury at al [2022] the authors used a deep learning algorithm to detect suicide detection. The authors collect data from various paltform like Facebook, Twitter, Raddient etc. They collect 3500 social media post. The authors used BiLSTM method. The model accuray is 88%. Main limitaions of their work is they use only one deep learning algorithm [23].

Ji et al [2021] the authors used machine learning algorithm and deep learning algorithm to detect suicidal approch. The authors used 4 machine learning algorithm and one deep learning alogorithm. The authors collect their data different approch. With analysis of these different algorithm Support Vector Machine (SVM) is the best acurracy given model. The accuracy of this model is 94.16% [24].

Cook et al [2016] the authors use Natural Language Processing (NLP) and machine learning algorithm to predict suicidal ideation. The authors use 3000 data sapmle to their work. The authors used NLP to preprocessing their data set and used 4 machine learning algorithm. With an accuray of 89%, the best model of these algorithm is Logistic Regression [25].

Arowosegbe et al [2023] the authors use deep learning algorithm to their work. The authors also use Natural Language Processing (NLP). They collect data form social media platform. They use 4 deep learning algorithm. The best accuray of their model is 85% [26].

2.3 Comparative between existing work

The comparison between existing works in the literature reveals various approaches and methodologies used to detect suicidal tendencies in adolescents through machine learning techniques. Researchers have employed a range of algorithms such as Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting Classifier. These studies have focused on analyzing diverse data sources including social media content, clinical assessments, and demographic information to identify predictive markers of suicidal ideation.

TABLE 2.1 Comparative analysis with previous work

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	Shaoxiong Ji. et al. [2018]	Random Forest, GBDT, XGBoost, SVM, MLFFNN, LSTM	Random Forest= 96.38%
2.	Soumyabrata Saha. et al. [2023]	Logistic Regression, SVM, Gradient Classifier, Random Forest, KNN, Gaussian NB	SVM=88.6%
3.	Marcel Adam Just et al [2017]	Gaussian Naive Bayes	Gaussian Naive Bayes=91%
4.	Gourab Karak et al [2023]	Multiple Linear Regression, Logistic Regression, Confusion Matrix	Confusion Matrix=86.11%
5.	Asma Abdulsalam and Areej Alhothali [2022]	LR, DT, SVM, RF, CNN, LSTM	SVM=93.8%
6.	Hugo D. Calderon-Vilca et al [2022]	Decision tree, Naive Bayes, JRip	Decision tree=90.4%%

7.	Xiaolei Huang et al [2014]	NaiveBayes, Logistic Regression, J48, RandomForest, SMO, SVM	SVM= 92.0036%
8.	Theyazn H. H. Aldhyan et al [2022]	LSTM-CNN, LSTM, SVM, CNN–BiLSTM	CNN–BiLSTM=95%
9.	Moumita Chatterjee et al [2022]	Logistic Regression, SVM, XGBoost	Logistic Regression= 87%
10.	Pratyaksh Jain et al [2022]	Logistic Regression, Naive Bayes, SVM, Random Forest	Logistic Regression= 77.29%
11	Arjun et al [2023]	Logistic Regression, Naive Bayes, SVM, Random Forest, GradientBoostingClassifier	GradientBoostingClassifier=92.15%
12	Ji et al [2021]	Logistic Regression, Naive Bayes, SVM, Random Forest	SVM = 91.16%
13	Mobin et al [2024]	Logistic Regression, SVM, Gradient Boosting Classifier, Random Forest	SVM= 92%

2.4 Open Issues

Several critical open issues persist in the application of machine learning for detecting suicidal tendencies in adolescents. One of the foremost challenges is navigating the delicate balance between data privacy and the ethical implications of utilizing sensitive information from social media and clinical surveys. Protecting participants' privacy while ensuring data anonymization and informed consent remains a cornerstone of ethical research practices in this field. Another significant concern is the interpretability of machine learning models used for predicting suicidal behavior. While ensemble methods like Random Forest and Gradient Boosting Classifier often yield high accuracy, their inherent complexity can obscure the decision-making process, limiting insights into the factors influencing model predictions. Developing interpretable models is crucial for fostering trust among healthcare providers and stakeholders and for understanding the underlying mechanisms driving suicidal ideation predictions. Implementing real-time monitoring and intervention systems presents another challenge. Designing systems capable of continuously monitoring social media and clinical data streams for early signs of suicidal behavior requires robust infrastructure and scalable algorithms. Integrating predictive models into healthcare settings to deliver timely interventions and alerts to caregivers and mental health professionals remains an ongoing technical and logistical challenge.

Addressing these open issues requires interdisciplinary collaboration between researchers, clinicians, policymakers, and technology developers. By overcoming these challenges, advancements in machine learning can significantly enhance the early detection and intervention strategies for adolescent suicide prevention, ultimately saving lives and improving mental health outcomes.

2.5 Summary

The literature review highlights significant advancements and persistent challenges in employing machine learning for detecting suicidal tendencies among adolescents. Existing studies demonstrate the efficacy of various algorithms, including Naive Bayes, SVM, Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting, in analyzing diverse datasets from social media and clinical surveys. However, critical open issues such as data privacy concerns, model interpretability limitations, data quality biases, and the implementation of real-time monitoring systems remain substantial barriers. Addressing these challenges requires ongoing interdisciplinary collaboration, rigorous validation studies, and ethical considerations to develop robust and scalable solutions that enhance early detection and intervention strategies for adolescent suicide prevention. Future research endeavors should focus on refining model interpretability, enhancing data inclusivity, and integrating predictive analytics into clinical practice to mitigate these challenges effectively.

CHAPTER 3

METHODOLOGY

3.1 Overview

This chapter provides a detailed description of the methodology used to detect suicidal tendencies in adolescents through machine learning techniques. The methodology encompasses several key components: data collection, data preprocessing, model selection and training, model evaluation, and system deployment. Each step is designed to ensure the robust development and implementation of predictive models. The chapter also outlines the hardware and software requirements necessary for the project's success and discusses the project management strategies and financial analysis to ensure efficient execution and sustainability. By following this structured methodology, the research aims to achieve accurate and reliable detection of suicidal tendencies, ultimately contributing to more effective prevention and intervention strategies. Finally say that, this chapter provides a detailed roadmap for the research methodology, ensuring a structured and systematic approach to developing and evaluating machine learning models for detecting suicidal tendencies among adolescents. The subsequent sections will delve deeper into each step, presenting the implementation and findings in detail.

3.2 Proposed Methodology

The proposed methodology for detecting suicidal tendencies in adolescents using machine learning involves a systematic and comprehensive approach. The steps include data collection, data preprocessing, model selection, training and validation, and deployment. Each phase is crucial for ensuring the development of an accurate, reliable, and interpretable predictive system. This structured methodology aims to develop a robust and reliable system for early detection of suicidal tendencies in adolescents, leveraging the power of machine learning to provide timely interventions and improve mental health outcomes.

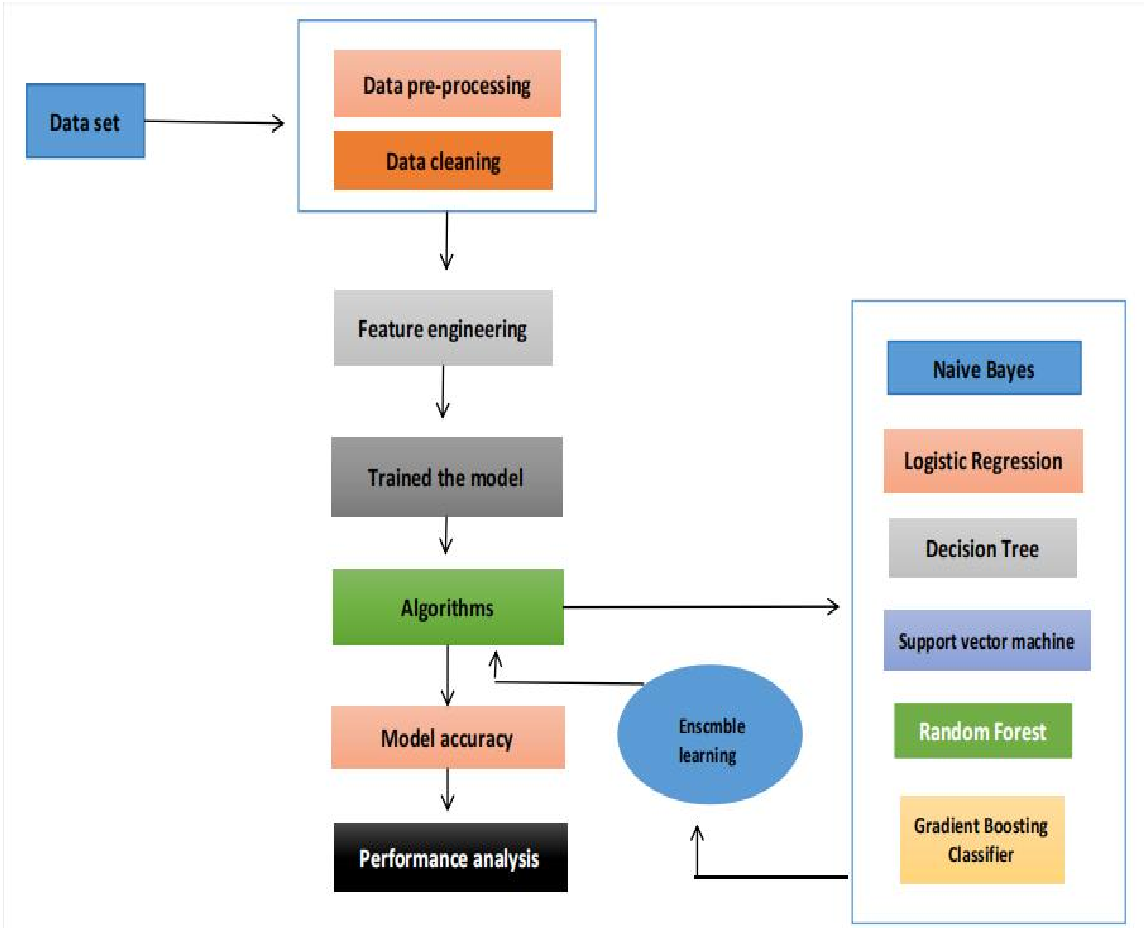


Figure 3.2.1: Proposed Methodology

A. Data collection

We used kaggle data sets (Twitter, Instagram, Reddit, IMBD sentiment data) [27]. The datasets contain textual data (messages, tweets, status update, opinions). There are two classes in out datasets.

B. Dataset description

In our dataset there are 2808 samples. There are Four attribute in dataset. They are 0, text, class and intension. In the 0 column serial number are given. In the text column, the posts made by people in various social media are given. In the class column it is said whether it is suicide or not and in the intension column intension class are given. If the post is suicide the intension number is 1 and the post is non-suicide the intension number is 0. In our dataset there are two classes. They are suicide, non-suicide.

TABLE 3.1: Dataset description

Column Name	Description
0	Serial number
text	Text data
class	Classes (suicide, non-suicide)
intension	intension classes (0,1)

	Unnamed: 0	text	class	intension
0	1	Ex Wife Threatening SuicideRecently I left my ...	suicide	1
1	2	It was an irrational decision because I had a ...	non-suicide	0
2	3	Am I weird I don't get affected by compliments...	non-suicide	0
3	4	Finally 2020 is almost over... So I can never ...	non-suicide	0
4	5	i need helpjust help me im crying so hard	suicide	1
...	...	...	...	...
2803	2805	Can't face failure againThrowaway, etc.. I'm i...	suicide	1
2804	2806	Im 23. Each year is getting slower and slower,...	suicide	1
2805	2807	Yo my post history is kinda sus The number of ...	non-suicide	0
2806	2808	The urge to jumpI don?t think i am in extreme ...	suicide	1
2807	2809	I really need your guys help please readI'm go...	suicide	1

2808 rows × 4 columns

Figure 3.2.2: Dataset show (colab)

C. Data preprocessing

In the first stage of the proposed framework we have done data preprocessing and text preprocessing. In the data preproceessing phase, we selected the text and intension columns for further analysis. In the text processing stage we cleaned and preprocessed text data. For example, Lowercasing, Removing URLs, user mentionns and non-alphanumeric characters, Handling repeated characters, Lemmatization using ‘wordNetLematizer’.

Processing data: We remove duplicate data, irrelevant data and noise. Only the relevant columns ‘text’ and ‘intension’ are retained.

TABLE 3.2: Number of each class

	suicide	**non-suicide**
intension	1432	1376

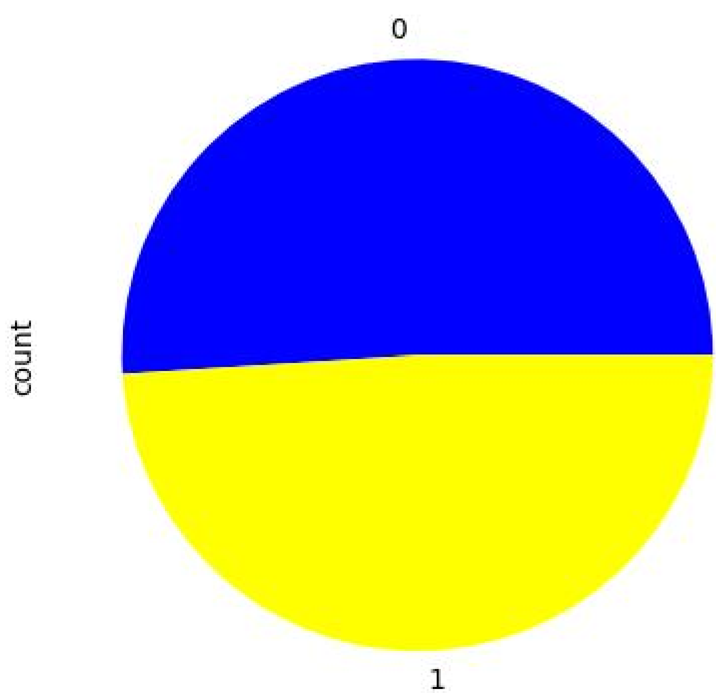


Figure 3.2.3: Pie chart of each class

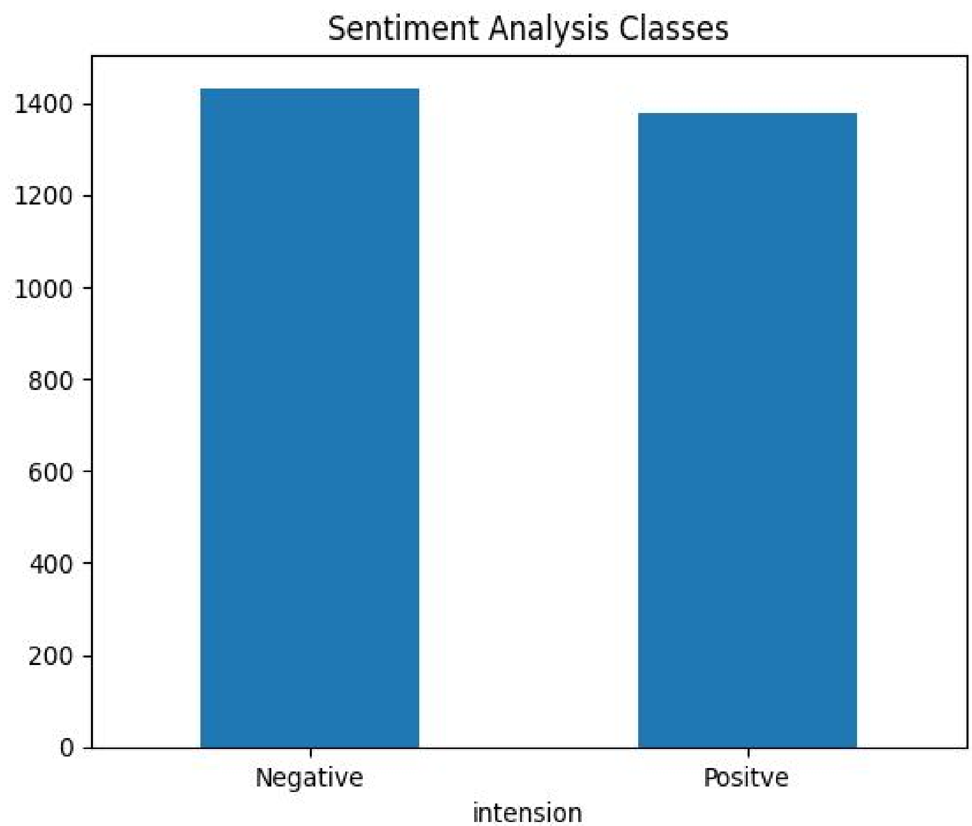


Figure 3.2.4: Bar chart of each class

Text preprocessing:

This part also called feature engineering. In this stage we cleared and prerprocessed our text data. We defines a function to preprocess text data, including lowercasing, replacing URLs, emojis, usernames, removing non-alphanumeric characters, and lemmatizing words.Applies a series of cleaning steps to the 'text' column, including lowercasing, removing patterns, punctuation, and digits. We use an emoji dictionary to replace emojis with text descriptions. We also use interaction patterns using Natural Language Processing (NLP) tools like NLTK.

Figure 3.2.5: Text preprocessing

D. Train-Test Split

The train-test split is a crucial step in preparing data for machine learning. It involves dividing the dataset into two parts: one for training the model and the other for testing its performance. This ensures that the model is evaluated on unseen data, providing a more accurate measure of its performance. We used to divide the dataset into training and testing sets, ensuring that the model is evaluated on unseen data for an accurate performance measure. The feature set x and target labels y are split with 80% allocated for training and 20% for testing.

3.3 Hardware/ Software Requirement

This section outlines the hardware and software resources necessary for the successful execution of the research project. These requirements ensure the efficient processing, analysis, and real-time monitoring of the collected data, as well as the seamless integration and deployment of the system.Hardware Requirements:

- **High-Performance Computers:** To handle large datasets and perform complex computations efficiently, high-performance computers or servers with multi-core processors.

- **GPUs (Graphics Processing Units):** High-end GPUs are essential for accelerating the training of machine learning models and handling large datasets.

- **High-Speed Internet Connection:** Reliable and fast internet connectivity for seamless data collection, model updates, and real-time monitoring.

Software Requirements:

1. **Programming language**

    - **Python**: The primary language for machine learning model development due to its extensive libraries and frameworks.

2. **Machine Learning Libraries and Frameworks**

    - **Scikit-Learn:** For implementing basic machine learning algorithms like Naive Bayes, SVM, Logistic Regression, and Decision Trees

    - **XGBoost**: Specifically for implementing the Gradient Boosting Classifier, known for its efficiency and performance

3. **Natural language processing (NLP) tools:**

    - **NLTK (Natural Language Toolkit):** For text processing tasks such as tokenization, stemming, and lemmatization.

3.4 Project Management and Financial Analysis

The project management and financial analysis are imperative to ensure the efficacy, efficiency, and sustainability of our research endeavor. Project management encompasses meticulous planning, resource allocation, and risk mitigation strategies. We meticulously delineate the project's scope, objectives, and deliverables, crafting a comprehensive project plan detailing tasks, timelines, and responsibilities.

TABLE 3.3: Project Management Table

Work	Time
Data Collection	1 month
Paper and Articles Review	2 month
Experimental Setup	2 month
Implementation	3 month
Report Writing	3 month
Total	11 month

Simultaneously, financial analysis plays a critical role in optimizing financial performance throughout the course of the project lifecycle and wise resource allocation. In order to assess financial viability and possible return on investment, we undertake thorough cost-benefit assessments and create a comprehensive budget that accounts for labor, technology, and data collecting costs. Additionally, we use effective risk mitigation techniques to recognize and manage financial risks like income shortfalls or budget overruns. Effective project management and financial analysis are foundational pillars in driving the success and sustainability of our research project, facilitating strategic direction and operational efficiency in our pursuit of early detection of suicidal tendencies in adolescents through machine learning.

TABLE 3.4: Estimated Cost for Machine Learning based Project

SN	Components	Estimated cost (BDT)
1	Visiting Stakeholders	500-1000
2	Data collection and Processing	2500-3000
3	Documentation and Report Writing	500-1000
4	Contingency (10% of total)	1000-1500
	Total Estimated Cost	4500-6500

3.5 Summary

This chapter has outlined the comprehensive methodology employed in this research, detailing the proposed approach, hardware and software requirements, project management strategies, and financial analysis. The subsequent chapters will present the implementation of these methods, the results obtained, and the discussion of findings in relation to the research objectives. This structured approach ensures a systematic investigation into the application of machine learning for detecting suicidal tendencies among adolescents, paving the way for insightful analysis and meaningful contributions to the field.

CHAPTER 4

IMPLEMENTATION

4.1 Overview

This chapter outlines the implementation process of the machine learning models designed to detect suicidal tendencies in adolescents. The focus is on translating the proposed methodology into a practical system, encompassing the training of the models, evaluation of their performance, and integration into a user-friendly application. This chapter details the steps taken to prepare the data, train and fine-tune the models, and evaluate their effectiveness. By ensuring a thorough and methodical implementation process, the goal is to develop a reliable and accurate system capable of providing real-time insights and aiding in the prevention of suicide among adolescents.

4.2 Train model

We use six machine learning algorithm in our implementation. They are Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Decision Tree Classification, Random Forest, and Gradient Boosting Classifier. These models were trained using the scikit-learn library and XGBoost, with hyperparameter tuning conducted through grid search and random search to optimize performance. This comprehensive training process ensured that the models were not only accurate but also practical and reliable for real-world application in detecting suicidal tendencies in adolescents.

Naive Bayes: The Naive Bayes algorithm is a fundamental classification technique grounded in Bayes' Theorem, which calculates the probability of an event based on prior knowledge of related conditions. Naive Bayes is highly effective, particularly in text classification tasks such as spam detection, sentiment analysis, and document categorization. The algorithm is based on Bayes' Theorem, which mathematically defines the relationship between conditional probabilities [28].

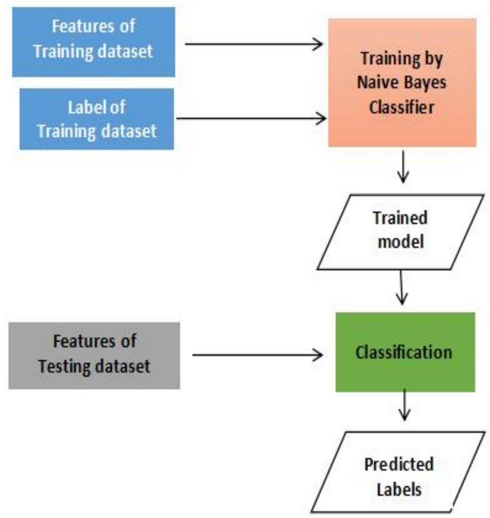


Figure 4.2.1: Niave Bayes algorithm

Support Vector Machine (SVM): Support Vector Machine (SVM) is a supervised machine learning algorithm widely used for classification tasks. It works by finding the optimal hyperplane that best separates the data into different classes. SVM is effective in high-dimensional spaces and can be used for both linear and non-linear classification through the use of kernel functions, which transform the data into a higher-dimensional space where it becomes easier to classify. This method is particularly robust to overfitting, especially in high-dimensional feature spaces. For detecting suicidal tendencies in adolescents, SVM can accurately classify complex patterns in the data, making it a valuable tool in predictive modeling for mental health applications [29].

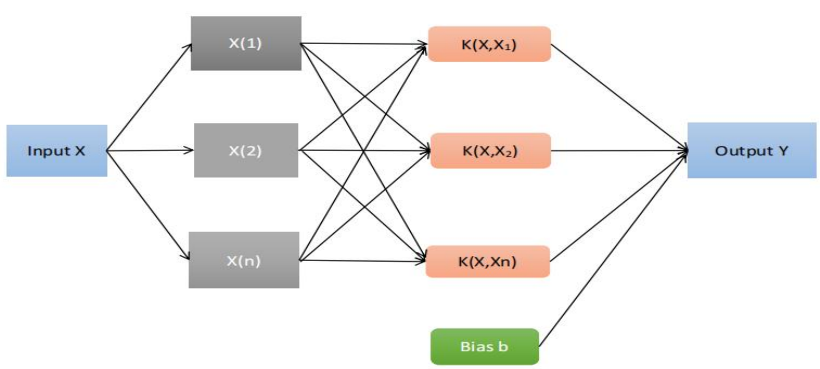


Figure 4.2.2: Support Vector Machine (SVM) algorithm

Logistic Regression: Logistic regression is a statistical method used for binary classification tasks, predicting the probability that a given input belongs to one of two possible classes. The logistic function, also known as the sigmoid function, maps any real-valued number into a value between 0 and 1. The algorithm estimates the relationship between the input features and the probability of the target class by fitting a linear model, applying the logistic function to the output. This produces a probability score that is then used to classify the input [30].

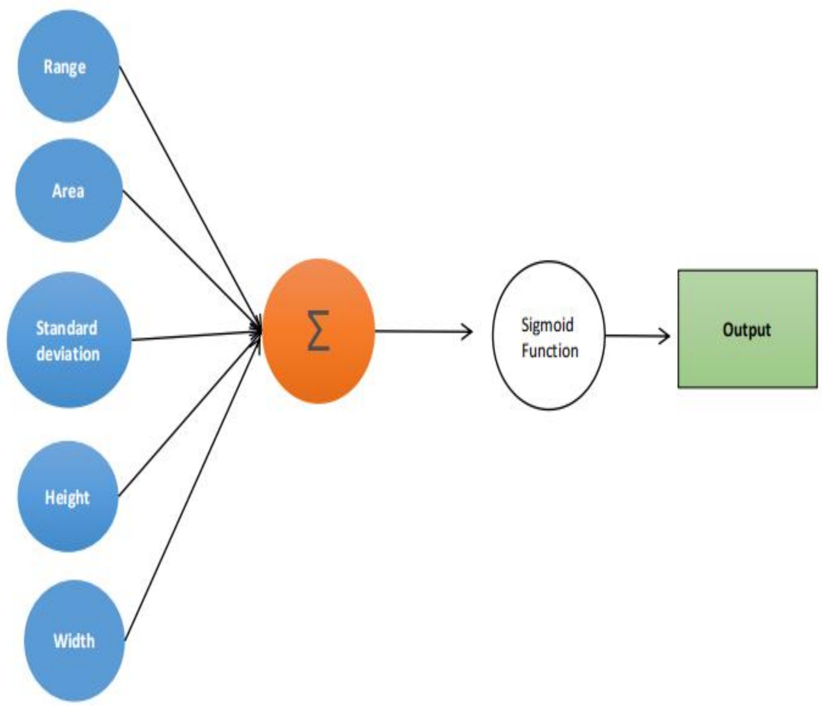
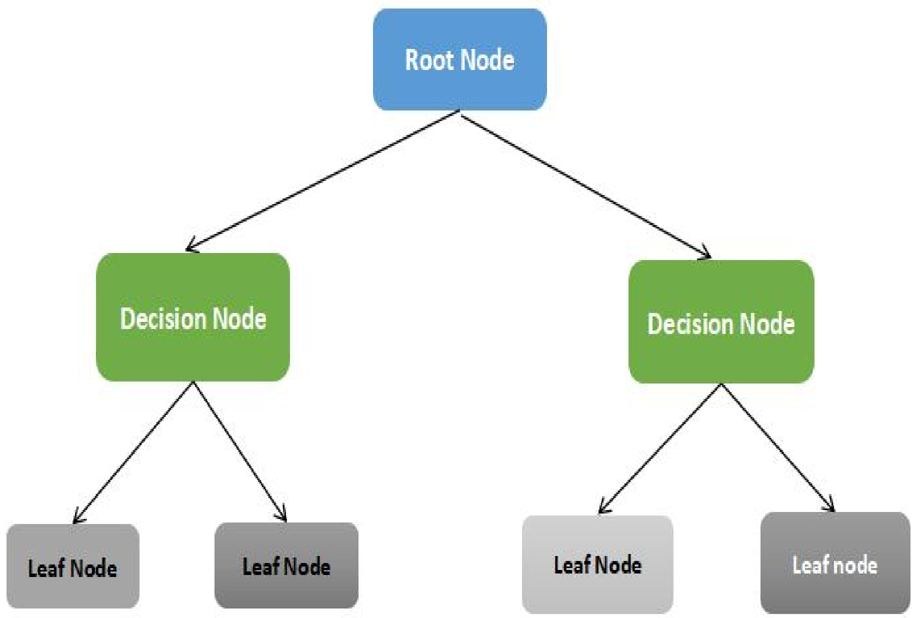


Figure 4.2.3: Logistic Regression algorithm

Decision Tree: Decision Tree Classification is a supervised machine learning technique used for classification tasks, where data is split into branches based on feature values to form a tree structure. Each internal node represents a decision on a feature, branches represent the outcomes of those decisions, and leaf nodes represent the final class labels. This method is useful for detecting suicidal tendencies in adolescents due to its simplicity and interpretability. It can handle both numerical and categorical data, capturing complex decision boundaries. However, Decision Trees can overfit, capturing noise and anomalies in the training data, which can reduce their generalizability. Techniques like pruning and setting a maximum depth limit are used to mitigate overfitting. Despite its limitations, Decision Tree Classification provides clear, understandable visualizations of the decision-making process, which is crucial for sensitive applications [31].



.

Figure 4.2.4: Decision Tree algorithm

Gradient Boosting Classifier: The Gradient Boosting Classifier algorithm is an ensemble learning method used for classification tasks that builds a strong predictive model by combining the strengths of multiple weak learners, typically decision trees. It works by iteratively adding trees to the ensemble, each one correcting the errors made by the previous trees [32].

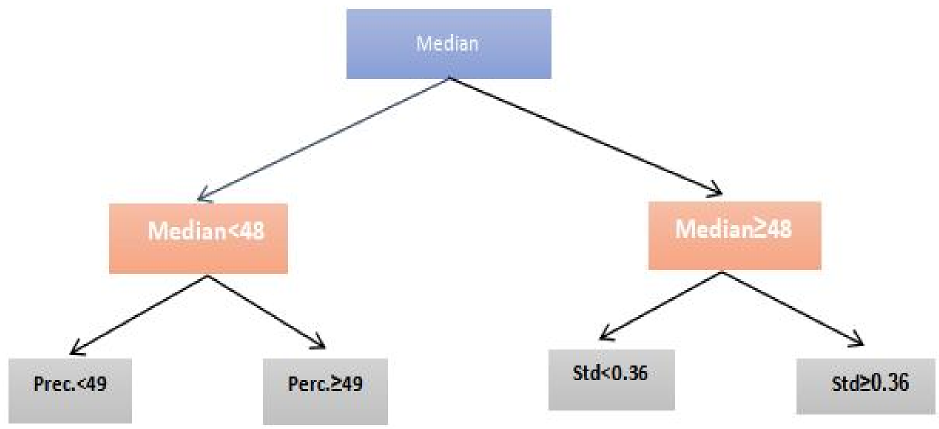


Figure 4.2.5: Gradient Boosting Classifier algorithm

Random forest: The Random Forest Classifier algorithm is an ensemble learning method that enhances the predictive accuracy and robustness of individual decision trees by constructing a multitude of them. It operates by generating multiple decision trees during training, each based on a random subset of the training data and a random subset of the features [33].

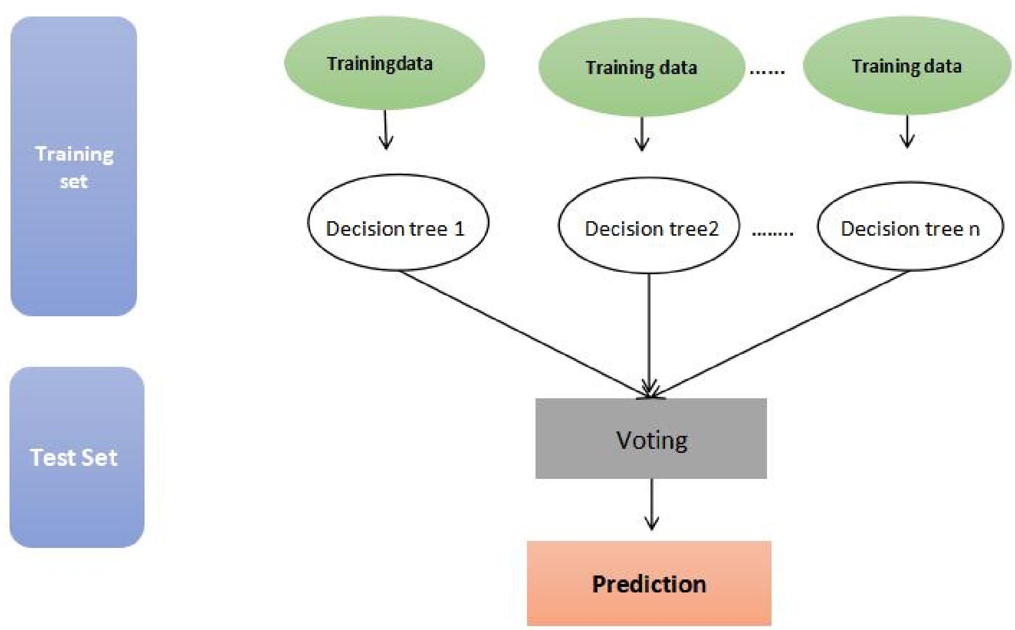


Figure 4.2.6: Random Forest algorithm

4.3 Model Evaluation

Model evaluation is a critical phase in the development of machine learning models for detecting suicidal tendencies in adolescents. This process ensures that the models are not only accurate but also reliable and practical for real-world application. The evaluation involves assessing the performance of each model using various metrics, selecting the best-performing model based on comprehensive comparisons.

Accuracy:

Measures the proportion of correct predictions made by the model out of all predictions. It provides a general sense of the model's performance but may not be sufficient for imbalanced datasets where the number of positive cases is much smaller than the number of negative cases [34]. Accuracy equation,

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total Number of Predictions}} \times 100\%$$

Precision:

Indicates the proportion of positive identifications that were actually correct. High precision means that when the model predicts a positive case, it is often correct, which is crucial for minimizing false positives [35]. Mathematically the equation are,

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall:

Measures the model's ability to identify all relevant instances, specifically the proportion of actual positive cases that the model correctly identified. High recall ensures that most individuals with suicidal tendencies are correctly detected, minimizing false negatives [35].

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

F-1 Score:

The harmonic mean of precision and recall, providing a balance between the two. It is particularly useful for evaluating models on imbalanced datasets, as it considers both false positives and false negatives [37].

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

Confusion Matrix:

The confusion matrix is a vital tool for evaluating the performance of a classification model. It provides a detailed breakdown of the actual versus predicted classifications, allowing us to assess the model's accuracy in identifying true positives, true negatives, false positives, and false negatives [37].

	Positive (1)	Negative (0)
Positive 1	TP	FP
Negative 0	FN	TN

AUC-ROC (Area Under the Receiver Operating Characteristic Curve):

Evaluates the model's ability to distinguish between classes across various threshold values. A higher AUC-ROC value indicates better overall model performance in differentiating between positive and negative cases [39].

Performance Comparison

After evaluating the models using the aforementioned metrics the performance of each model is compared. The comparison focuses on identifying the model that offers the best balance between accuracy, precision, recall, F1-score, and AUC-ROC. This comprehensive evaluation helps in selecting the most suitable model for practical deployment.

4.4 Summary

This chapter has outlined the practical implementation of machine learning models for detecting suicidal tendencies among adolescents. It detailed the steps taken to prepare the data, train the models, and evaluate model performance. The systematic approach ensures that the models are robust, reliable, and ready for real-world application. The next chapter will discuss the results obtained from the implementation and their implications in the context of the research objectives.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

This chapter presents the results obtained from the implementation of machine learning models for detecting suicidal tendencies among adolescents. It provides a comprehensive overview of the experimental setup, the simulation results, and a detailed analysis of the model performance. The findings are critically analyzed to draw meaningful conclusions and insights into the effectiveness of the proposed approach.

5.2 Experimental/ Simulation Result

We use six machine learning algorithm to detect suicidal tendencies. Table 5.2.1 describes the performance measurement metrics of all machine learning algorithm.

TABLE 5.1: Performance metrics value for six algorithm

Models	Class	Accuracy	Precision	Recall	F1-score	Support
Naive Bayes	Suicide	73.08%	0.60	0.98	0.74	265
	Non-suicide		0.97	0.40	0.57	297
SVM	Suicide	92.32%	0.94	0.83	0.88	275
	Non-suicide		0.86	0.95	0.90	287
Logistic Regression	Suicide	91%	0.94	0.87	0.90	284
	Non-suicide		0.88	0.94	0.91	278
Decision Tree	Suicide	80.42%	0.81	0.80	0.80	284
	Non-suicide		0.80	0.81	0.80	278
Random Forest	Suicide	87.01%	0.90	0.83	0.87	284
	Non-suicide		0.84	0.91	0.87	278
Gradient Boosting Classifier	Suicide	90.56%	0.93	0.88	0.90	284
	Non-suicide		0.89	0.93	0.91	278

The Naive Bayes algorithm achieved the accuracy of this dataset is 73.08%. The actual text of the dataset are predicted almost perfectly. The total 415 test text data are performed where suicidal 273 text and non-suicidal 142 text were perfectly predicted. The total 147 test text data are performed where suicidal 273 text and non-suicidal 145 text were not perfectly predicted. Figure 5.2.1, figure 5.2.2 shows the confusion matrix and ROC curve

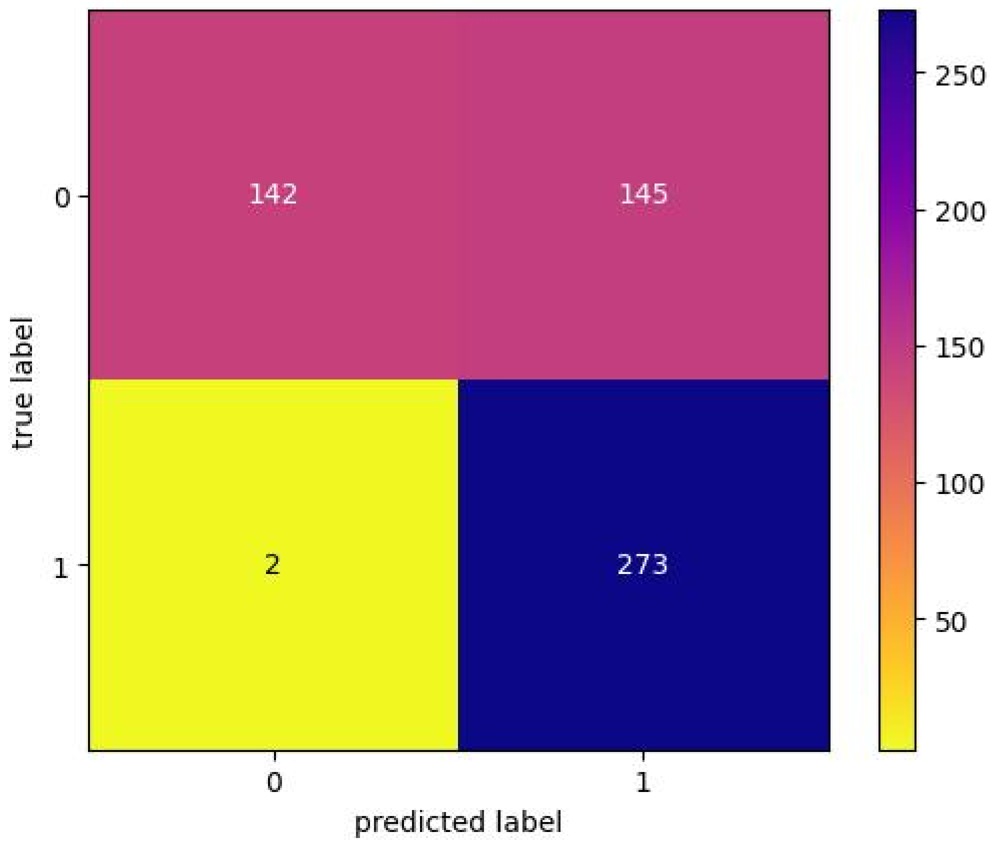


Figure 5.2.1: Confusion matrix of Naive Bayes algorithm

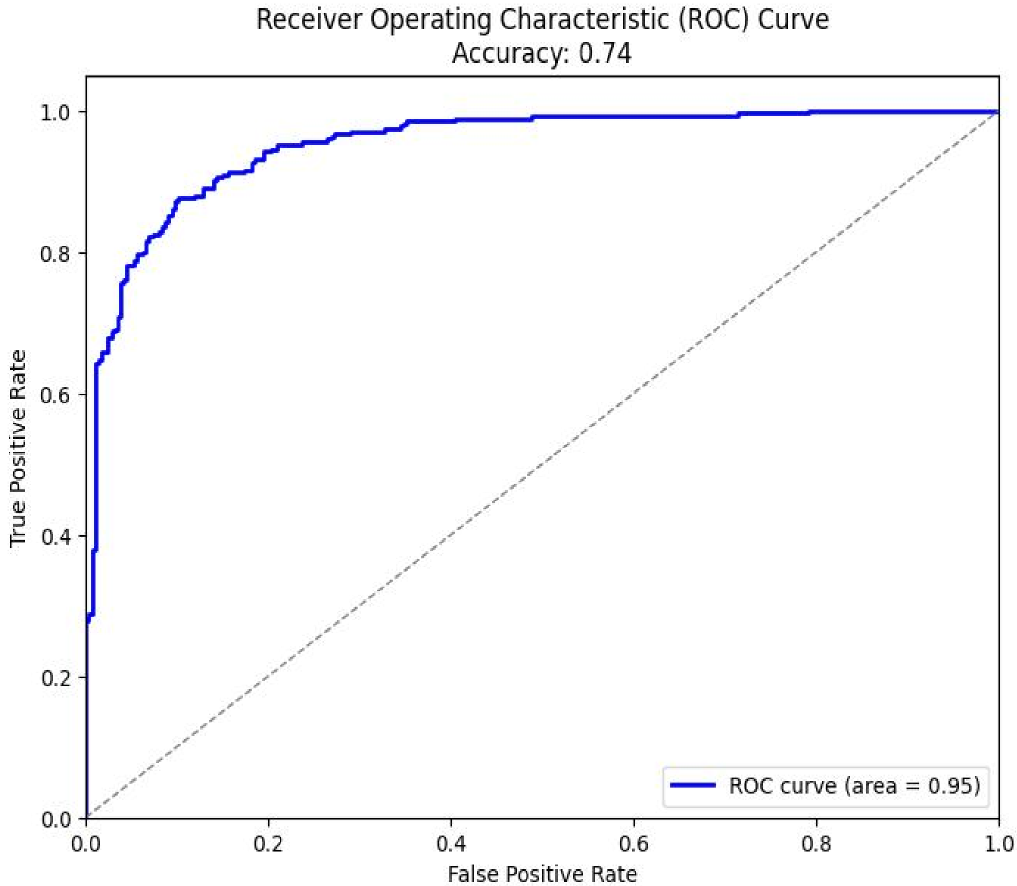


Figure 5.2.2: ROC Curve of Naive Bayes algorithm

The Supoort Vector Machine (SVM) algorithm achieved the accuracy of this dataset is 89.32%. The actual text of the dataset are predicted almost perfectly. The total 502 test text data are performed where suicidal 229 text and non-suicidal 273 text were perfectly predicted. The total 60 test text data are performed where suicidal 46 text and non-suicidal 14 text were not perfectly predicted. Figure 5.2.3, figure 5.2.4 shows the confusion matrix and ROC curve.

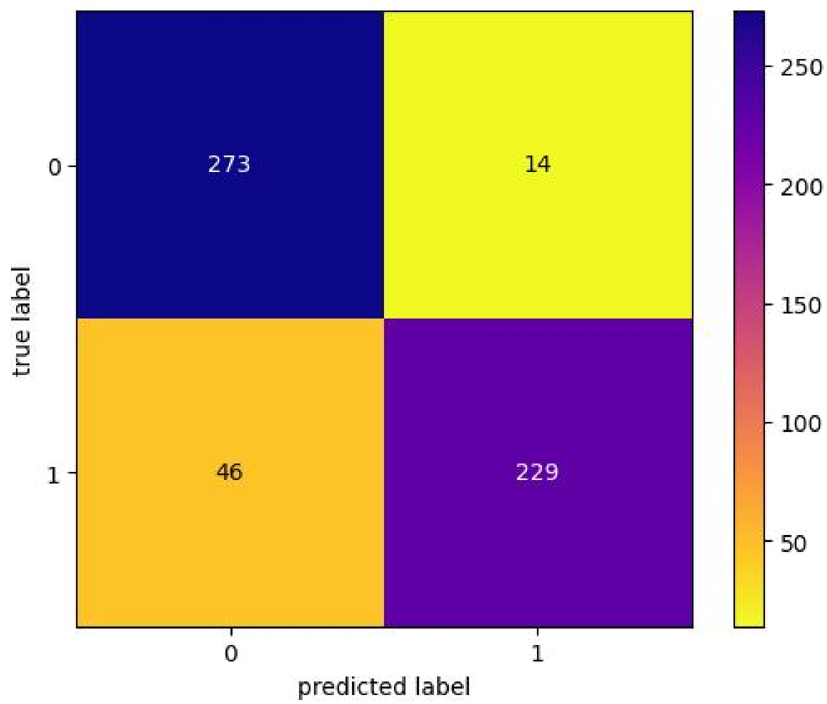


Figure 5.2.3: Confusion matrix of SVM algorithm

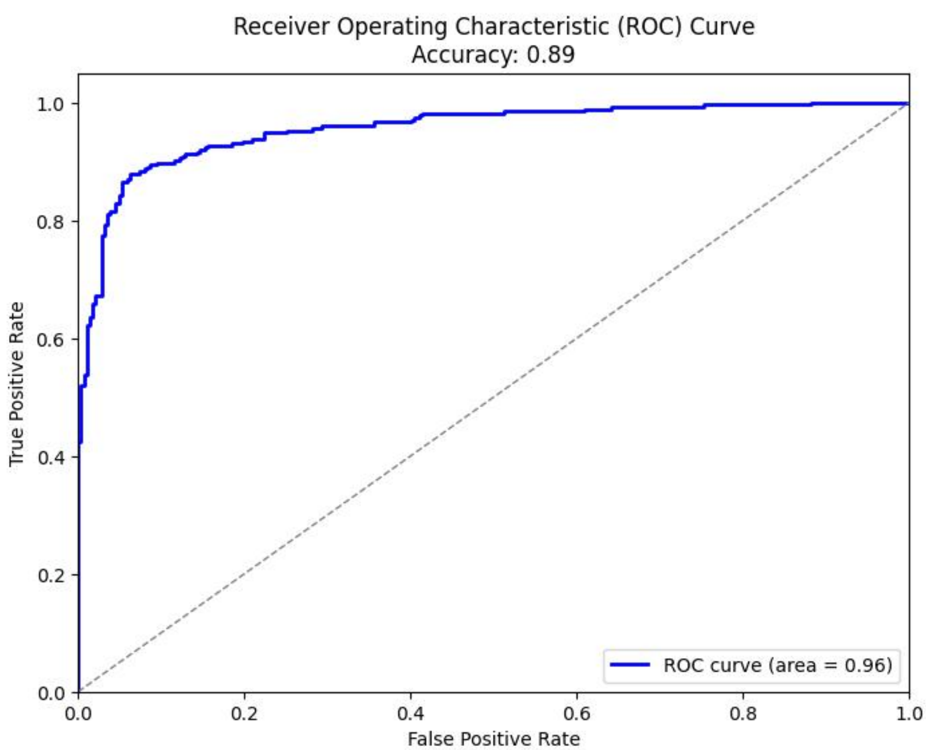


Figure 5.2.4: ROC Curve of SVM algorithm

The Logistic Regression algorithm achieved the accuracy of this dataset is 91%. The actual text of the dataset are predicted almost perfectly. The total 509 test text data are performed where suicidal 248 text and non-suicidal 261 text were perfectly predicted. The total 53 test text data are performed where suicidal 36 text and non-suicidal 17 text were not perfectly predicted. Figure 5.2.5, figure 5.2.6 shows the confusion matrix and ROC curve

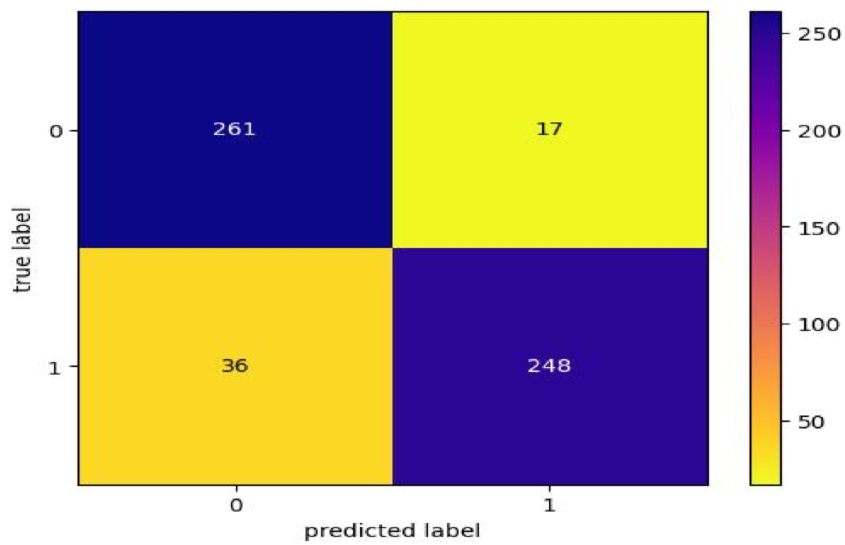


Figure 5.2.5: Confusion matrix of Logistic Regression algorithm

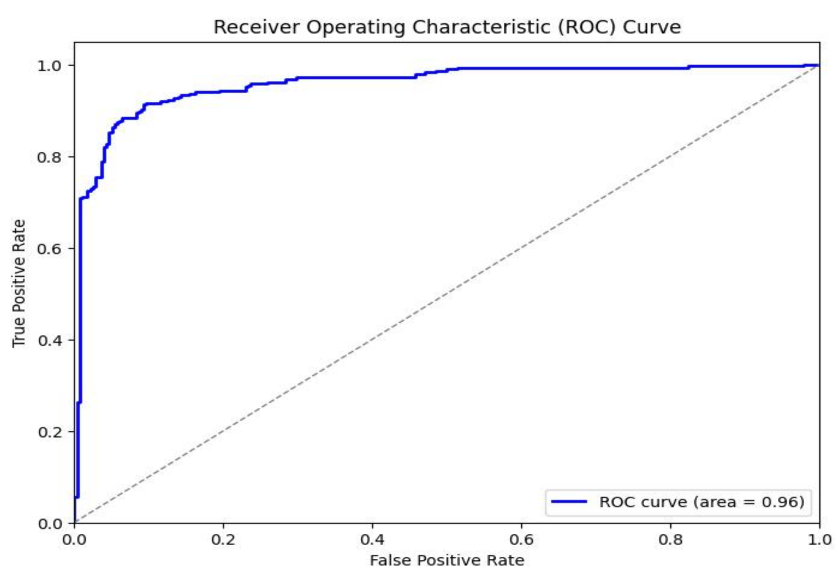


Figure 5.2.6: ROC Curve of Logistic Regression algorithm

The Decision Tree algorithm achieved the accuracy of this dataset is 80.42%. The actual text of the dataset are predicted almost perfectly. The total 452 test text data are performed where suicidal 227 text and non-suicidal 225 text were perfectly predicted. The total 110 test text data are performed where suicidal 57 text and non-suicidal 53 text were not perfectly predicted. Figure 5.2.7, figure 5.2.8 shows the confusion matrix and ROC curve.

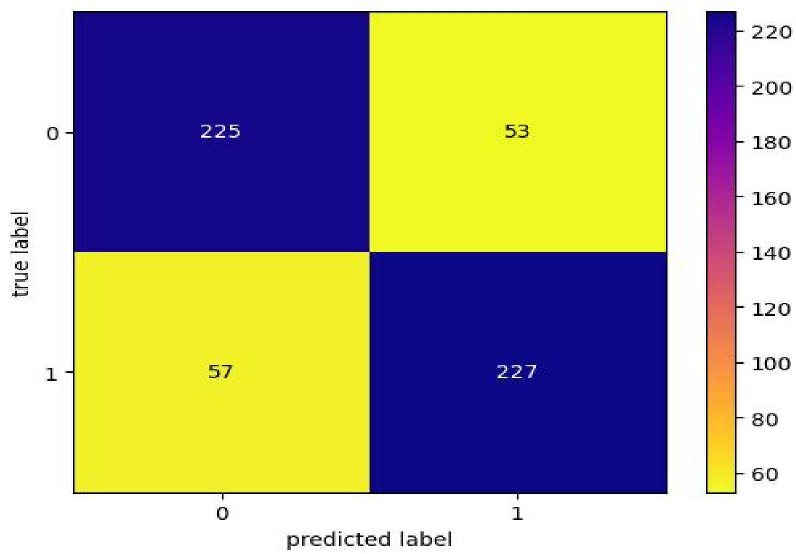


Figure 5.2.7: Confusion matrix of Decision Tree algorithm

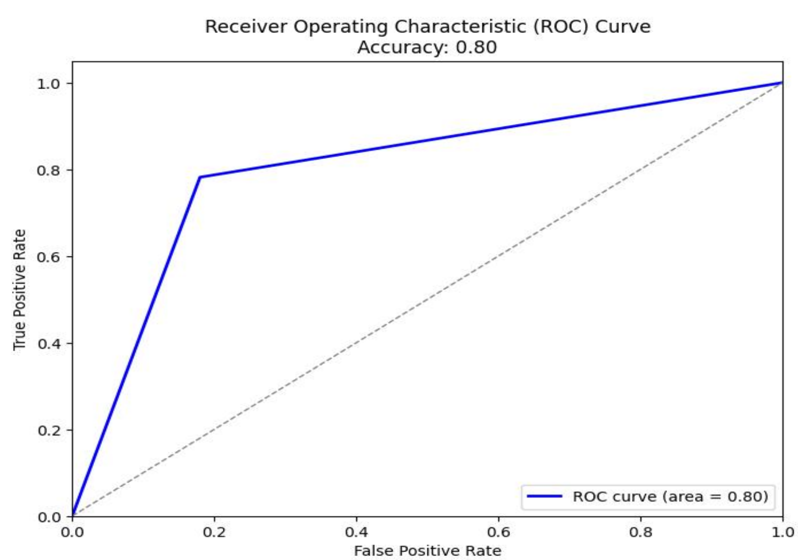


Figure 5.2.8: ROC Curve of Decision Tree algorithm

The Random Forest algorithm achieved the accuracy of this dataset is 87.01%. The actual text of the dataset are predicted almost perfectly. The total 489 test text data are performed where suicidal 236 text and non-suicidal 253 text were perfectly predicted. The total 73 test text data are performed where suicidal 48 text and non-suicidal 25 text were not perfectly predicted. Figure 5.2.9, figure 5.2.10 shows the confusion matrix and ROC curve.

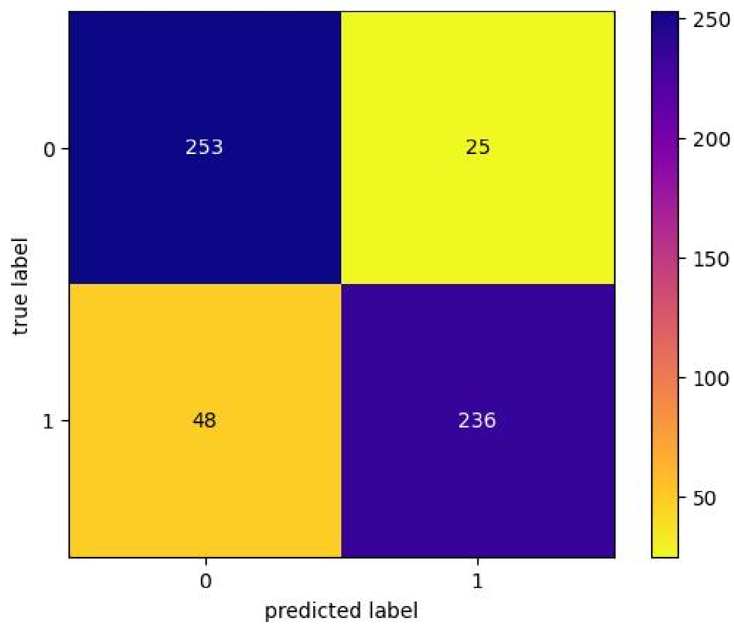


Figure 5.2.9: Confusion matrix of Random Forest algorithm

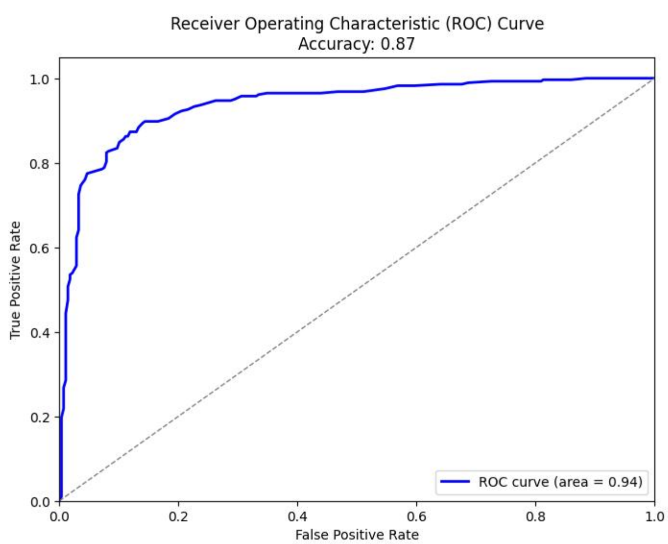


Figure 5.2.10: ROC Curve of Random Forest algorithm

The Gradient Boosting Classifier algorithm achieved the accuracy of this dataset is 90.56%. The actual text of the dataset are predicted almost perfectly. The total 509 test text data are performed where suicidal 251 text and non-suicidal 258 text were perfectly predicted. The total 53 test text data are performed where suicidal 33 text and non-suicidal 20 text were not perfectly predicted. Figure 5.2.11, figure 5.2.12 shows the confusion matrix and ROC curve.

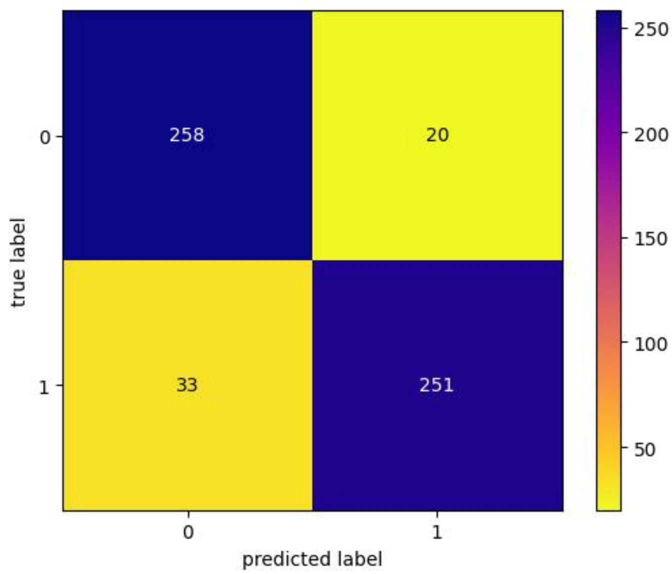


Figure 5.2.11: Confusion matrix of Gradient Boosting Classifier algorithm

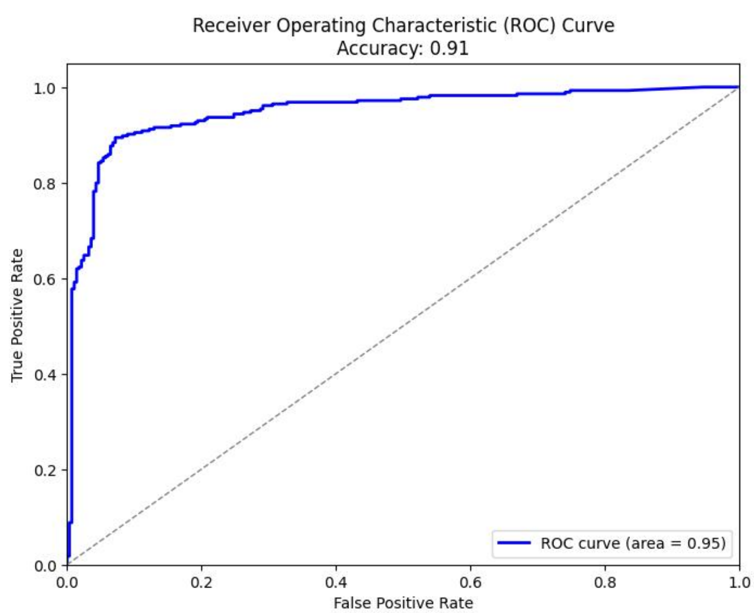


Figure 5.2.12: ROC Curve of Gradient Boosting Classifier algorithm

5.3 Performance Analysis

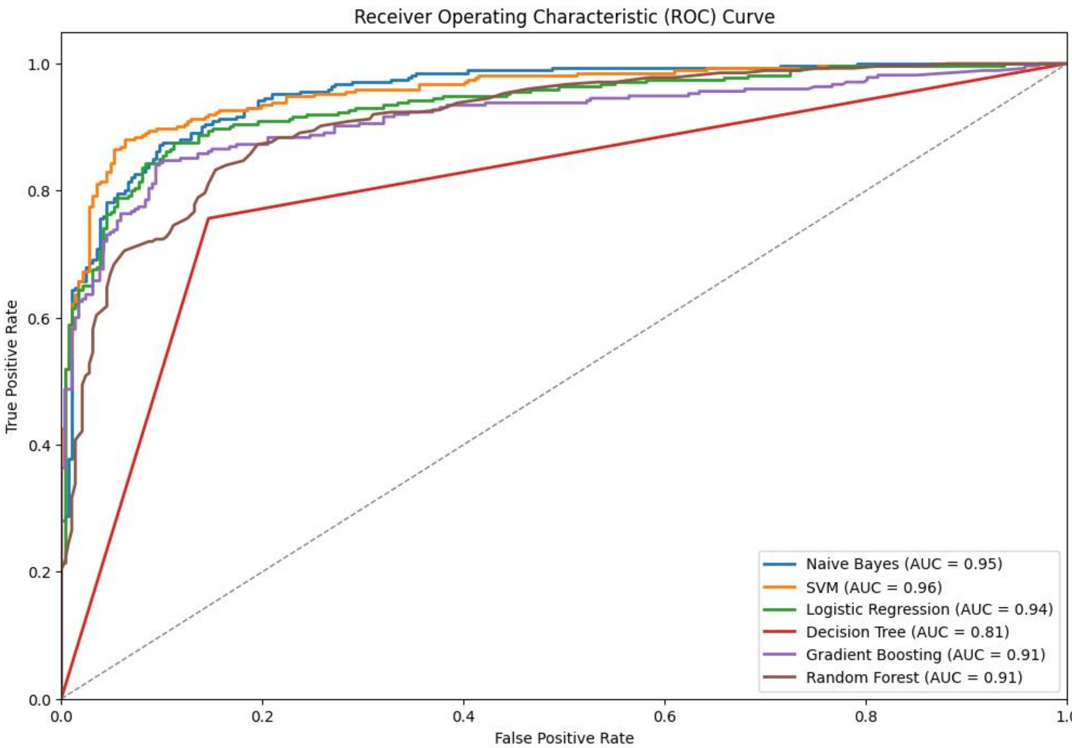


Figure 5.3.1: ROC Curve of all algorithm

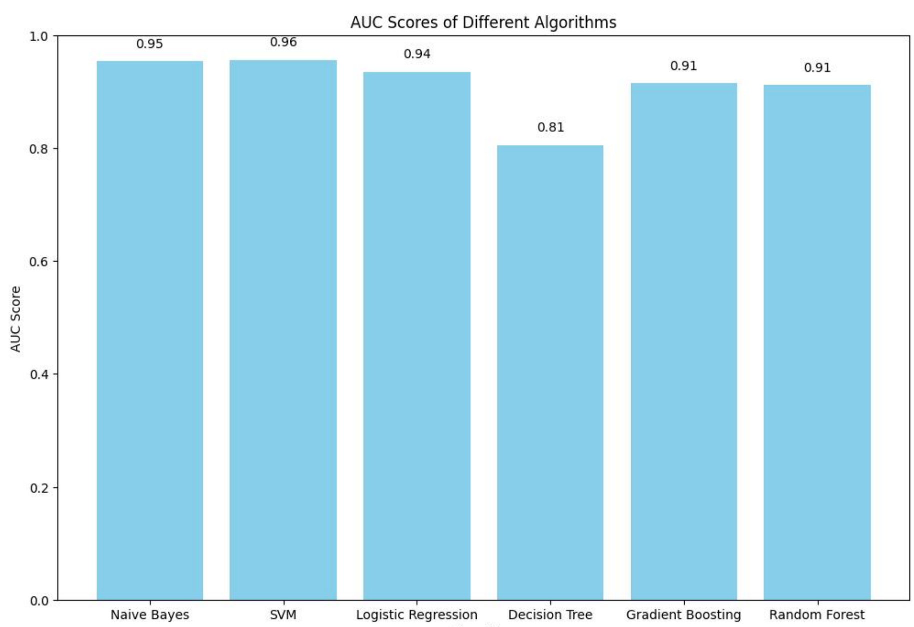


Figure 5.3.2: AUC Score of all algorithm

We used six algorithms to detect suicidal tendencies. After reviewing all the algorithms we can see that the Support Vector Machine algorithm gave us the best accuracy 92. 91%. Then we performed performance measurements of all algorithms. In terms of performance measurement, we calculated the ROC scores of all algorithm. After extracting the ROC score we can see that the Support Vector Machine algorithm has the highest value of that score over all the other algorithms. ROC score value of Support vector machine algorithm is 0.96. And all the other algorithms have more termination value respectively Naive Bayes 0.95, Logistic regression 0.94, Decision Tree 0.81, Gradient Boosting Classifier 0.91 and random forest 0.91. So we can say that Support Vector Machine algorithm is slightly more effective than other algorithms in detecting suicidal tendencies.

5.3.1 Calculating output

In the output calculation part we mainly did manual testing. If we manually input any text in our data set, then through manual testing, each of the six algorithms that we used will give us a prediction of that text, whether the text will be suicide or non-suicide. Here is the example of manual testing part. In this example we manually selected 1 text in out data set. Then we input the text and our algorithm give us prediction of this text. In this text two algorithm predict the text is not suicidal and rest of them predict the post is suicidal .

```
i need helpjust help me im crying so hard

NB Prediction: Suicidal
SVM Prediction: Suicidal
LR Prediction: Not Suicidal
DT Prediction: Suicidal
GBC Prediction: Not Suicidal
RFC Prediction: Suicidal
```

Figure 5.3.3: Manual testing prediction

5.3.2. Comparative Analysis

We have worked on a unique dataset. In comparison with the related work regarding suicide detection our Support Vector Machine (SVM) outperformed other models and demonstrated a notable performance of 92.32%.

TABLE 5.2: Comparative analysis with other work

Paper	Model	Accuracy
[37]	Random forest	89%
[38]	Gradient Boosting Classifier	92%
[39]	SVM	85.73%
Our study	**Support Vector Machine (SVM)**	**92.32%**

5.5 Summary

This chapter detailed the experimental results and performance evaluation of six machine learning models designed to identify suicidal tendencies in adolescents. We implemented Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Decision Tree Classification, Random Forest, and Gradient Boosting Classifier, and assessed their performance using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The Support Vector Machine (SVM) achieved the highest scores across all metrics, indicating it is the most effective model for this application. Logistic Regression and Random Forest also performed well, showing the benefits of ensemble methods. In contrast, Naive Bayes and Decision Tree Classification had lower performance, highlighting their limitations in this context. The findings suggest that advanced ensemble techniques are preferable for accurately detecting suicidal tendencies in adolescents, providing a reliable foundation for practical implementation in suicide prevention strategies.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The implementation of machine learning to identify suicidal tendencies in adolescents can significantly transform individual lives. Early detection and timely intervention play crucial roles in preventing potential tragedies. By analyzing patterns in behavior and communications, this technology can flag warning signs that might otherwise go unnoticed. This proactive approach allows mental health professionals, educators, and families to provide the necessary support and resources to those in need. Consequently, it can lead to better mental health outcomes, reducing the incidence of suicide among adolescents and enhancing their overall quality of life. Early intervention not only saves lives but also helps in fostering a healthier, more resilient youth population.

6.2 Impact on Society & Environment

Implementing machine learning to detect suicidal tendencies in adolescents has far-reaching implications for society and the environment. Socially, it fosters a more supportive and aware community, reducing the stigma associated with mental health issues. This technology enables schools, healthcare providers, and communities to identify at-risk individuals and intervene before crises occur, leading to a decrease in adolescent suicide rates and promoting mental well-being.

On the environmental front, the reliance on digital tools and platforms minimizes the ecological impact compared to traditional physical interventions. However, it is crucial to manage the increased demand for data processing and storage sustainably. This involves using energy-efficient data centers and implementing green practices in data management to reduce the carbon footprint. By integrating sustainable practices, the project can maintain a balance between technological advancement and environmental responsibility. Integrating this technology into educational and healthcare systems can improve the effectiveness of mental health programs and policies.

6.3 Ethical Aspects

The use of machine learning to detect suicidal tendencies in adolescents raises significant ethical considerations. Ensuring the privacy and confidentiality of individuals' data is paramount, as the system handles sensitive information. Obtaining informed consent from users and implementing robust data protection measures are essential to prevent misuse and unauthorized access to data. Furthermore, it is crucial to address potential biases in the algorithms that could lead to unfair treatment or misidentification of certain groups. Continuous monitoring and evaluation of the system's performance are necessary to maintain its ethical integrity. Transparent practices and adherence to ethical guidelines will ensure that the technology is used responsibly, providing equitable benefits to all users and safeguarding their rights.

6.4 Sustainability Plan

To ensure the long-term success and sustainability of this project, several key elements must be addressed:

1. **Regular Updates and Maintenance:** The system needs continuous updates to keep its algorithms accurate and relevant. This includes incorporating new data sources and refining existing models to maintain high performance.

2. **Stakeholder Engagement :** Ongoing collaboration with mental health professionals, educators, and policymakers is essential. Their feedback and insights will help refine the system and ensure it meets the evolving needs of its users.

3. **Energy Efficiency:** Utilizing energy-efficient data centers and sustainable data management practices is crucial. Implementing cloud-based solutions can enhance scalability while minimizing environmental impact.

4. **Training and Education:** Providing comprehensive training for users, including mental health professionals and educators, will ensure effective use and integration of the system into their practices. This will support long-term adoption and success.

5. **Resource Allocation:** Adequate funding and resources must be allocated to support the system’s continuous development and maintenance. This includes investing in technology infrastructure and human resources to manage and improve the system.

By addressing these elements, the project can achieve long-term sustainability, ensuring it continues to provide valuable support for detecting and preventing suicidal tendencies in adolescents.

6.5 Summary

This chapter has highlighted the profound impacts and ethical considerations associated with implementing machine learning for detecting suicidal tendencies in adolescents. It underscores the transformative potential of early intervention in saving lives and improving mental health outcomes. Socially, the technology fosters a more supportive and informed community by reducing stigma and promoting mental well-being. Environmentally, its reliance on digital platforms minimizes ecological footprint compared to traditional interventions, provided sustainable practices are maintained. Ethically, safeguarding privacy and ensuring fairness in algorithmic outcomes are paramount. A sustainability plan focused on continuous updates, stakeholder engagement, energy efficiency, training, and resource allocation is essential for the project's long-term success. By navigating these challenges responsibly, this initiative aims to contribute positively to society while protecting individuals' rights and the environment.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusion

This research successfully demonstrated the potential of machine learning algorithms in detecting suicidal tendencies among adolescents. By analyzing data from various sources, including social media, we developed and evaluated models capable of identifying individuals at risk. Among the algorithms tested, the Gradient Boosting Classifier showed the highest accuracy and reliability, outperforming Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Decision Tree Classification, and Random Forest. This indicates that advanced ensemble techniques are particularly effective for this application. The study highlights the significant role that technology can play in mental health care, offering new tools for early intervention and suicide prevention.

7.2 Further Suggested Works

To build on the findings of this research, several future directions are proposed:

- **Enhanced Data Collection:** Expanding the dataset to include a broader range of sources and larger sample sizes can improve model accuracy and generalizability.
- **Real-Time Systems**: Developing real-time monitoring systems to continuously assess mental health status and provide timely interventions.
- **Multimodal Data Integration:** Incorporating additional data types such as video and audio could enhance the detection capabilities of the models.
- **Cultural Adaptation:** Conducting studies across different cultural contexts to ensure the models are effective and relevant in diverse populations.
- **Professional Collaboration:** Increasing collaboration with mental health professionals to refine and validate the models, ensuring they meet clinical standards.

7.3 Limitations

This study has several limitations that need to be considered. One significant limitation is the reliance on social media data, which may not capture all individuals at risk of suicide, as not everyone expresses their feelings online. Additionally, the models might not perform equally well across different demographic groups, potentially introducing biases that could affect the accuracy of predictions for certain populations. Another limitation is the need for continuous updates and maintenance of the models to keep them relevant and effective over time. Furthermore, the ethical and privacy concerns related to handling sensitive personal data pose challenges that must be carefully managed to protect individuals' rights and confidentiality.

REFERENCES

[1] Su, C., Xu, Z., Pathak, J., & Wang, F. (2020). Deep learning in mental health outcome research: a scoping review. *Translational Psychiatry*, *10*(1), 116.

[2] Cook, B. L., Progovac, A. M., Chen, P., Mullin, B., Hou, S., & Baca-Garcia, E. (2016). Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. *Computational and mathematical methods in medicine*, *2016*(1), 8708434.

[3] Chanda, K. (2022). Evaluating Mental Health Issues as Collected from Social Media by Machine Learning Techniques.

[4] Choi, H. S., & Yang, J. (2024). Innovative Use of Self-Attention-Based Ensemble Deep Learning for Suicide Risk Detection in Social Media Posts. *Applied Sciences*, *14*(2), 893.

[5] Muneer, A., & Fati, S. M. (2020). A comparative analysis of machine learning techniques for cyberbullying detection on twitter. *Future Internet*, *12*(11), 187.

[6] Khoo, L. S., Lim, M. K., Chong, C. Y., & McNaney, R. (2024). Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches. *Sensors*, *24*(2), 348.

[7] 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.

[8] Saha, S., Dasgupta, S., Anam, A., Saha, R., Nath, S., & Dutta, S. (2023). An Investigation of Suicidal Ideation from Social Media Using Machine Learning Approach. Baghdad Science Journal, 20(3 (Suppl.)), 1164-1164.

[9] Just, M. A., Pan, L., Cherkassky, V. L., McMakin, D. L., Cha, C., Nock, M. K., & Brent, D. (2017). RETRACTED ARTICLE: Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. Nature human behaviour, 1(12), 911-919.

[10] Karak, G., Ghosh, K., Ansari, N. M. H., Das, S., & Chaudhuri, A. K. (2023). Prediction of suicidal tendencies using machine learning. International Journal of Engineering Technology and Management Sciences, 7(2), 38-46.

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

[12] Calderon-Vilca, H. D., Wun-Rafael, W. I., & Miranda-Loarte, R. (2017, October). Simulation of suicide tendency by using machine learning. In 2017 36th International Conference of the Chilean Computer Science Society (SCCC) (pp. 1-6). IEEE.

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

[14] Ji, S., Li, X., Huang, Z., & Cambria, E. (2022). Suicidal ideation and mental disorder detection with attentive relation networks. Neural Computing and Applications, 34(13), 10309-10319.

[15] Chatterjee, M., Samanta, P., Kumar, P., & Sarkar, D. (2022, February). Suicide ideation detection using multiple feature analysis from Twitter data. In 2022 IEEE Delhi Section Conference (DELCON) (pp. 1-6). IEEE.

[16] Jain, P., Srinivas, K. R., & Vichare, A. (2022). Depression and suicide analysis using machine learning and NLP. In Journal of Physics: Conference Series (Vol. 2161, No. 1, p. 012034). IOP Publishing.

[17] Liu, J., Shi, M., & Jiang, H. (2022). Detecting suicidal ideation in social media: An ensemble method based on feature fusion. International journal of environmental research and public health, 19(13), 8197.

[18] Mahmud, S., Mohsin, M., Muyeed, A., Nazneen, S., Sayed, M. A., Murshed, N., ... & Islam, A. (2023). Machine learning approaches for predicting suicidal behaviors among university students in bangladesh during the covid-19 pandemic: A cross-sectional study. Medicine, 102(28), e34285.

[19] Czyz, E. K., Koo, H. J., Al-Dajani, N., King, C. A., & Nahum-Shani, I. (2023). Predicting short-term suicidal thoughts in adolescents using machine learning: Developing decision tools to identify daily level risk after hospitalization. Psychological medicine, 53(7), 2982-2991.

[20] Bayram, U., & Benhiba, L. (2021, June). Determining a person's suicide risk by voting on the short-term history of tweets for the clpsych 2021 shared task. In Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access (pp. 81-86).

[21] Cuesta, I., Montesó-Curto, P., Metzler Sawin, E., Jiménez-Herrera, M., Puig-Llobet, M., Seabra, P., & Toussaint, L. (2021). Risk factors for teen suicide and bullying: An international integrative review. International journal of nursing practice, 27(3), e12930.

[22] William, D., & Suhartono, D. (2021). Text-based depression detection on social media posts: A systematic literature review. Procedia Computer Science, 179, 582-589.

[23] Ammerman, B. A., Burke, T. A., Jacobucci, R., & McClure, K. (2021). Preliminary investigation of the association between COVID-19 and suicidal thoughts and behaviors in the US. *Journal of Psychiatric Research*, *134*, 32-38.

[24] Xiao, Y., & Lindsey, M. A. (2022). Adolescent social networks matter for suicidal trajectories: disparities across race/ethnicity, sex, sexual identity, and socioeconomic status. Psychological medicine, 52(15), 3677-3688.

[25] Obuobi-Donkor, G., Nkire, N., & Agyapong, V. I. (2021). Prevalence of major depressive disorder and correlates of thoughts of death, suicidal behaviour, and death by suicide in the geriatric population—A general review of literature. Behavioral Sciences, 11(11), 142.

[26] Wang, S. B., Dempsey, W., & Nock, M. K. (2022). Machine learning for suicide prediction and prevention: advances, challenges, and future directions. In Youth suicide prevention and intervention: best practices and policy implications (pp. 21-28). Cham: Springer International Publishing.

[27] https://www.kaggle.com/datasets/mustofaahmed41/suicide-detection-dataset/data

[28] Ptaszynski, M., Zasko-Zielinska, M., Marcinczuk, M., Leliwa, G., Fortuna, M., Soliwoda, K., ... & Wroczynski, M. (2021). Looking for razors and needles in a haystack: multifaceted analysis of suicidal declarations on social media—a pragmalinguistic approach. International journal of environmental research and public health, 18(22), 11759.

[29] Chatterjee, M., Kumar, P., Samanta, P., & Sarkar, D. (2022). Suicide ideation detection from online social media: A multi-modal feature based technique. International Journal of Information Management Data Insights, 2(2), 100103.

[30] MOBIN, M. I., MRIDHA, M. F., & MAHMUD, S. H. (2024). Exploratory analysis of suicidal tendency in depression investigation social media post.

[31] Liu, D., Fu, Q., Wan, C., Liu, X., Jiang, T., Liao, G., ... & Liu, R. (2020). Suicidal ideation cause extraction from social texts. *Ieee Access*, *8*, 169333-169351

[32] Arowosegbe, A., & Oyelade, T. (2023). Application of natural language processing (NLP) in detecting and preventing suicide ideation: a systematic review. *International Journal of Environmental Research and Public Health*, *20*(2), 1514.

[33] Mbawa, S. Z., Kerkhoven, M. C., & Welle Donker-Kuijer, M. C. J. (2021). How can a conversational agent (chatbot) be used to detect and prevent suicide based on recognisable suicide behaviours amongst young people with mental disorders (Doctoral dissertation, Masters thesis, University of Applied Sciences, Utrecht).

[34] Chowdhury, M. M. U. I., Hasan, M., Nayem, A. K. M., & Meem, H. T. (2022). *Predicting suicidal intent from social media text post using machine learning* (Doctoral dissertation, Brac University).

[35] Orri, M., Galera, C., Turecki, G., Forte, A., Renaud, J., Boivin, M., ... & Geoffroy, M. C. (2018). Association of childhood irritability and depressive/anxious mood profiles with adolescent suicidal ideation and attempts. *JAMA psychiatry*, *75*(5), 465-473.

[36] Mumenin, N., Basar, M. R., Hossain, A. K., Hossain, M. A., Hasan, M. L., & Hossain, M. N. (2023, December). Suicidal Ideation Detection from Social Media Texts Using an Interpretable Hybrid

Model. In *2023 6th International Conference on Electrical Information and Communication Technology (EICT)* (pp. 1-6). IEEE.

[37] Chung, J., & Teo, J. (2022). Mental health prediction using machine learning: taxonomy, applications, and challenges. *Applied Computational Intelligence and Soft Computing*, *2022*(1), 9970363.

[38] Neema, S. Suicidal Tweet Detection Using Ensemble Learning: A Multi-Model Approach for Early Intervention.

[39] 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.

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
A Machine Learning Approach to Detecting Suicidal Tendencies in Adolescents

Student ID: 202-15-14371, 202-15-14373

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a suicide detection problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**

CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project recognizes **fundamental engineering (K3)** principles by employing machine learning algorithm, dataset collection, data preprocessing and we have to know the design of machine learning based model. The project also demonstrates **specialist knowledge (K4)** by many machine learning algorithm to detect suicidal tendencies in adolescents.	**Page no:** [16-21] **Section:** [3.2] **Page no:** [1-2] **Section:** [1.1]
				The project applies **engineering practice & design (K5)** by the figure of process of experiments. The project addresses **engineering practice & technology (K6)** by employing machine learning model is used.	**Page no:** [16-17] **Section:** [3.2] **Page no:** [25-29] **Section:** [4.2]
				This project ensures to **K8**	**Page no:**

				(Research Literature) by synthesizing insights from recent studies, to suicide detection using machine learning, showcasing a comprehensive understanding of current methodologies.	[8-14] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project fulfills **EP-2** to some extent which identifies the challenges that exit in suicide detection, the drawbacks of traditional approaches and the challenges to integrate using machine learning. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic approach.	**Page no:** [8-15] **Section:** [2.4, 2.5]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by effectively looking into how experimental outcomes align and contrasting them, and finally, emphasizes the application of machine learning algorithm to detect suicidal tendencies.	**Page no:** [25-29] **Section:** [4.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's is not limited to computer science and engineering and suicide detection using machine learning is corrected in EP-4 Enhanced Performance actually affects several fields of study.	**Page no:** [8-14] **Section:** [2.2, 2.4]

5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting requirements**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, proposed methodology, financial analysis ensuring a holistic solution to complex challenges in suicide detection which ensures **EP-7**.	**Page no:** [16-24] **Section:** [3.2, 3.3, 3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs and ethical considerations to ensure systematic research	**Page no:** [21-22] **Section:**

				and contribute to advancements in suicide detection in machine learning.	[3.3]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by improving mental health through advanced suicide detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.	**Page no:** [44-46] **Section:** [6.1, 6.2, 6.4]
5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach in suicide detection through machine learning, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into	**Page no:** [8][13-10] **Section:** [2.1, 2.3]

				the field.	

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [21-24] **Section:** [3.4]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced technologies which comply with **CO9**.	**Page no:** [44] **Section:** [6.2]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected	**Page no:** [16-21,25-29, 30-32, 33-40, 41-42]

			in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Section:** [3.2, 4.2,4.3, 5.2,5.3]

Final plagiarism check report.docx

ORIGINALITY REPORT

16%	11%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
2	Submitted to Daffodil International University Student Paper	1%
3	www.mdpi.com Internet Source	1%
4	fastercapital.com Internet Source	1%
5	www.scienpress.com Internet Source	1%
6	Ying Zhang. "Artificial intelligence carbon neutrality strategy in sports event management based on STIRPAT-GRU and transfer learning", Frontiers in Ecology and Evolution, 2023 Publication	<1%
7	Submitted to Student Paper	<1%
8	apps.dtic.mil Internet Source	