DEPRESSION DETECTION FROM SOCIAL MEDIA ACTIVITY USING MACHINE LEARNING

BY

BIPASHA BABUL BRISHTI
ID: 0242310005103031

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

Supervised By

Dr. Arif Mahmud
Associate Professor & Program Director MIS
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Abdus Sattar
Assistant Professor & Coordinator M.sc
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2024

APPROVAL

This Project/Thesis titled **Depression Detection From Social Media Activity Using Machine Learning** , submitted by Bipasha Babul Brishti, ID No: 0242310005103031 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 M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 06-07-2024.

BOARD OF EXAMINERS

Chairman

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

Internal Examiner

Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Ohidujjaman, PhD
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Mr. Sadat Hossain
Data Scientist
Risk Management Division,
BRAC Bank Limited

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. Arif Mahmud Associate Professor & Program Director MIS Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Dr. Arif Mahmud
Associate Professor & Program Director MIS
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Abdus Satter
Assistant Professor & Coordinator M.sc
Department of CSE
Daffodil International University

Submitted by:

Brishti

Bipasha Babul Brishti
ID: **0242310005103031**
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I 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 our profound our indebtedness to Dr. Arif Mahmud Associate Professor & Program Director MIS 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 draft and correcting them at all stage have made it possible to complete this project.

I would like to express our heartiest gratitude to Dr.Sheak Rashed Haider Noori 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

In my research project titled "Depression detection from social media activity using machine learning," I investigate the effectiveness of various machine learning algorithms in identifying depressive symptoms from social media data. Throughout the study, I evaluate the performance of Support Vector Machine (SVM), Random Forest, Long Short-Term Memory (LSTM), and Bidirectional LSTM (Bi-LSTM) algorithms. Utilizing a dataset comprising labeled social media posts, I analyze the results obtained from each algorithm. The SVM algorithm achieves an accuracy of 90%, with precision and recall scores of 0.87 and 0.94 for depressed posts, and 0.94 and 0.86 for non-depressed posts, respectively. Meanwhile, Random Forest yields an accuracy of 79%, while LSTM and Bi-LSTM models perform significantly better, achieving accuracies of 97% and 98%, respectively. These deep learning models demonstrate superior performance, as evidenced by their precision, recall, and F1-score metrics. My findings underscore the potential of machine learning, particularly deep learning techniques, in revolutionizing depression detection and mental health monitoring through the analysis of social media activity.

TABLE OF CONTENTS

LIST OF FIGURE

CHAPTER 1
INTRODUCTION

1.1 Introduction

In this digital age, social media platforms have become integral to our daily lives, this offering unprecedented opportunity for communication, expression, and community building. However, these platforms also provide a unique lens through which the mental health of individuals can be observed and analyzed. Depression that a pervasive mental health disorder, affects millions worldwide and is often reflected in the language, tone, and engagement patterns of users on social media. This project, "Depression Detection from Social Media Activity Using Machine Learning," aims to harness the power of machine learning algorithms to identify signs of depression in social media posts [1]. By leveraging natural language processing (NLP) techniques and sentiment analysis, this study seeks to develop a robust model capable of detecting depressive tendencies from textual data. The motivation behind this research stems from the increasing rates of depression globally and the urgent need for early detection and intervention. Traditional methods of diagnosing depression are often resource-intensive and require clinical settings, which can be barriers for many individuals. In contrast, a machine learning-based approach can provide a scalable and non-intrusive means of screening large populations, potentially identifying at-risk individuals who might otherwise remain undiagnosed. The project involves collecting a diverse dataset of social media posts, preprocessing the data to ensure quality and relevance, and training various machine learning models to classify posts as indicative or non-indicative of depression. Key challenges addressed include handling the vast and noisy nature of social media data, dealing with privacy and ethical considerations, and ensuring the model's accuracy and generalizability across different demographics and platforms. The outcomes of this research could significantly impact mental health care by providing a tool for early detection, enabling timely support and intervention, and ultimately contributing to better mental health outcomes. This project not only showcases the potential of artificial intelligence in addressing pressing societal issues but also highlights the intersection of technology, psychology, and ethics in the modern digital landscape [2].

1.2 Motivation

The motivation behind my research project, "Depression Detection from Social Media Activity Using Machine Learning," is rooted in the profound impact of depression on individuals and society, coupled with the transformative potential of technology in addressing mental health challenges. Depression is a pervasive mental health disorder that characterized by persistent feelings of sadness, loss of interest in activities, and a variety of physical and emotional problems. It affects millions of people globally, contributing to significant personal distress, impaired social and occupational functioning, and even increased risk of suicide. The traditional methods of diagnosing depression often rely on self-reported symptoms and clinical evaluations, which can be time-consuming, stigmatizing, and inaccessible to many individuals due to various barriers, including lack of awareness, limited availability of mental health services, and societal stigma associated with mental health issues [3].

In parallel, the advent of social media has revolutionized people communicate and express themselves. Platforms such as Twitter, Facebook, and Instagram have become digital arenas where individuals share their thoughts, emotions, and experiences openly and frequently. This vast repository of user-generated content presents a unique opportunity to observe and analyze patterns of behavior and language that may be indicative of mental health conditions, including depression. The hypothesis driving this project is those individuals experiencing depression exhibit discernible patterns in their social media activity, such as changes in language use, sentiment, engagement levels, and interaction patterns. Using leveraging machine learning algorithms and natural language processing (NLP) techniques, it is possible to develop models capable of detecting these subtle cues and identifying users who may be at risk of depression.

The primary motivation for this research is to provide a scalable, non-intrusive, and timely method for early detection of depression. Early identification of depressive symptoms is crucial for effective intervention and treatment, potentially preventing the escalation of the disorder and reducing the associated personal and societal burdens. Machine learning-based approaches offer the advantage of processing vast amounts of data rapidly and continuously, enabling real-time monitoring and analysis of social media activity. This capability can complement traditional mental health services by providing an additional layer of screening and support, especially in under-resourced settings where access to mental health professionals is limited. Furthermore, the increasing prevalence of depression among various demographics, particularly adolescents and

young adults, underscores the urgency of developing innovative solutions for mental health monitoring. Social media is especially prevalent among these groups, making it a valuable data source for detecting mental health issues. By focusing on social media activity, this project aims to reach a population that may be more reluctant to seek help through conventional channels due to stigma or lack of awareness [4].

Another critical aspect of the motivation for this project is the potential to advance the field of human mental health research. The integration of machine learning and NLP in mental health assessment represents a cutting-edge approach that can yield new insights into the linguistic and behavioral markers of depression. This interdisciplinary research not only contributes to the development of practical tools for mental health professionals but also enriches our understanding of the complex interplay between digital behavior and psychological well-being.

Moreover, the ethical implications of this research cannot be overlooked. Ensuring the privacy and confidentiality of social media users is paramount, and the project is designed with strict adherence to ethical guidelines and data protection standards. The goal is to create a system that respects user privacy while providing valuable insights for mental health care. This ethical framework is essential for gaining public trust and fostering the responsible use of technology in sensitive areas such as mental health.

The motivation for this project is multifaceted, encompassing the urgent need for early detection of depression, the transformative potential of machine learning and NLP in mental health monitoring, the opportunity to address the mental health needs of vulnerable populations, and the ethical considerations inherent in leveraging social media data. By developing a machine learning-based system for detecting depression from social media activity, this project aims to contribute to the improvement of mental health care, the advancement of research, and the well-being of individuals and society at large. The successful implementation of this project holds the promise of making a significant positive impact on how we understand, detect, and address depression in the digital age [6].

1.3 Rationale of the Study

Rationale behind this study, "Depression Detection from Social Media Activity Using Machine Learning," stems from the growing recognition of depression as a critical public health issue and the unique potential of digital technology to address it. Depression is a leading cause of disability

worldwide, affecting individuals' that quality of life, productivity, and overall well-being. Despite the prevalence of depression, many individuals do not receive timely diagnosis or treatment due to a variety of barriers, including stigma, lack of access to mental health services, and underreporting of symptoms. Traditional diagnostic methods often rely on self-reported measures and clinical interviews, which can be subjective and resource-intensive. This underscores the need for innovative, scalable approaches to mental health assessment and intervention [7].

Social media platforms have become an integral part of modern life, with billions of users sharing their thoughts, feelings, and experiences online. These platforms provide a rich source of data that reflects users' mental states and behaviors in real-time. The hypothesis that underpins this study is that individuals experiencing depression exhibit distinct patterns in their social media activity, such as changes in language use, sentiment, and engagement levels. By analyzing these patterns using advanced level of machine learning algorithms and natural language processing (NLP) techniques, it is possible to detect signs of depression with high accuracy. The rationale for utilizing machine learning in this context is multifaceted. Firstly, machine learning algorithms excel at identifying complex patterns and correlations within large datasets, making them well-suited for analyzing the vast amounts of unstructured data generated on social media. These algorithms can be trained to recognize linguistic and behavioral markers of depression, enabling automated and continuous monitoring of users' mental health. This approach offers a significant advantage over traditional methods by providing a scalable solution that can process data in real-time, potentially identifying at-risk individuals before their condition worsens.

Secondly, the accessibility and ubiquity of social media make it a valuable tool for reaching populations that might otherwise be underserved by conventional mental health services. Adolescents and young adults, in particular, are prolific users of social media and are also among the most vulnerable to mental health issues. By leveraging social media data, this study aims to develop a detection system that can serve as an early warning mechanism, prompting timely interventions and support for those in need. This proactive approach could bridge the gap between the onset of depressive symptoms and the receipt of professional help, ultimately improving mental health outcomes [8].

The integration of machine learning and NLP in mental health assessment represents a cutting-edge advancement in the field of mental health research. This interdisciplinary study not only contributes to the development of practical tools for mental health professionals but also enhances

our understanding of the linguistic and behavioral manifestations of depression. The insights gained from this research can inform the creation of more effective and personalized interventions, tailored to the specific needs and characteristics of different user groups.

Ethical considerations are a critical component of this study's rationale. Ensuring the privacy and confidentiality of social media users is paramount, and the research design includes stringent measures to protect user data. The study adheres to ethical guidelines and data protection standards, aiming to balance the benefits of early depression detection with the imperative to respect individuals' privacy rights. This ethical framework is essential for gaining public trust and fostering the responsible use of technology in mental health applications.

The rationale for this study is grounded in the urgent need to address the global burden of depression through innovative and scalable solutions. By harnessing the power of machine learning and the rich data available on social media, this study aims to develop a reliable and efficient system for detecting depression early. This approach has the potential to transform mental health care by providing timely support to those in need, advancing our understanding of depression, and promoting the ethical use of technology in health monitoring. The successful implementation of this project could lead to significant improvements in how we diagnose, understand, and treat depression, ultimately contributing to better mental health outcomes and enhanced quality of life for individuals worldwide [9].

1.4 Research Questions

The research questions guiding this project, "Depression Detection from Social Media Activity Using Machine Learning," are designed to address the multifaceted challenges and opportunities presented by the intersection of mental health assessment and advanced technological methodologies.

The primary research question is:

1. How can machine learning algorithms be effectively utilized to detect depression from social media activity?
2. What are the key linguistic and behavioral markers in social media activity that correlate with clinical symptoms of depression?
3. What role does natural language processing play in enhancing the accuracy and reliability of depression detection models from social media data?

4. What is the potential impact of early detection of depression through social media activity on mental health outcomes, including timely intervention and access to mental health services?

1.5 Expected Output

The expected output of my research project, "Depression Detection from Social Media Activity Using Machine Learning," encompasses both the development of a robust, scalable system and the generation of valuable insights into the relationship between social media activity and mental health. Primarily, the project aims to deliver a machine learning-based model capable of accurately detecting signs of depression from social media posts. This model will be trained using a comprehensive dataset of social media activity, incorporating natural language processing (NLP) techniques to analyze text for linguistic and behavioral markers indicative of depression. The system is expected to classify social media posts with high accuracy, precision, and recall, thereby identifying users who may be experiencing depressive symptoms. Additionally, the project will provide a detailed evaluation of various machine learning algorithms and their effectiveness in this context [10]. By comparing the performance of different models, such as support vector machines, neural networks, and decision trees, the research will identify the most effective approaches for depression detection from social media data. This comparative analysis will be documented in the project report, offering insights into the strengths and limitations of each algorithm and guiding future research in this domain. Another significant output is the creation of a comprehensive set of linguistic and behavioral markers that correlate with depression. These markers, identified through the application of NLP techniques, will be thoroughly documented and analyzed, contributing to the understanding of how depression manifests in social media activity. This knowledge can inform the development of more targeted and effective mental health interventions and support tools. The project will also address the ethical considerations involved in using social media data for mental health assessment. The report will outline the strategies implemented to ensure user privacy and consent, providing a framework for ethical data handling in similar future projects. This ethical framework is a crucial output, as it ensures the responsible use of technology and builds public trust in such initiatives.

The expected output of this project includes a highly accurate machine learning model for depression detection, a comprehensive evaluation of different algorithms, a detailed set of

depression-related markers, an ethical framework for data use, and insights into the practical impact of early detection. These outcomes will not only advance the field of mental health assessment but also provide a foundation for future research and development in using technology to address mental health challenges. The successful implementation and documentation of these outputs will demonstrate the feasibility and value of integrating machine learning with social media analysis to detect and address depression in a scalable and ethical manner [11].

1.6 Project Management and Finance

Project Management Approach:

- Scope Definition: Clearly define the objectives, scope, and deliverables of the project, including the development of a machine learning model for depression detection.
- Resource Identification: Identify the necessary resources, including human resources, tools, and datasets required for the project.
- Timeline Establishment: Establish a detailed timeline with milestones and deadlines for each phase of the project, such as data collection, model development, testing, and deployment.
- ● Risk management: Recognize possible hazards, such as algorithmic biases, data privacy issues, and technological difficulties, and devise mitigation techniques.
- Communication Plan: Implement a communication plan to ensure regular updates and collaboration among team members, stakeholders, and external partners.
- Progress Monitoring: Follow up on project developments by using key performance indicators (KPIs) and making necessary plan adjustments to stay on schedule.

Financial Planning:

- Budget Allocation: Allocate funds for various project activities, including data acquisition, software development, personnel salaries, and overhead costs.
- Cost Estimation: Estimate costs for hardware, software licenses, cloud computing services, and any other expenses associated with the project.
- Funding Sources: Identify potential funding sources, such as grants, sponsorships, or partnerships with research institutions or industry partners.
- Return on Investment (ROI) Analysis: Assess the potential ROI of the project, considering the social and economic benefits of early depression detection and intervention.

- Financial Reporting: Maintain accurate financial records and prepare regular reports to track expenses, monitor budget adherence, and ensure accountability.

1.7 Report Layout

The report for the project "Depression detection from social media activity using machine learning" is structured as follows: Chapter 1 provides an introduction, discussing the motivation behind the study, the rationale, research questions, expected outputs, project management, financial aspects, and an overview of the report layout. Chapter 2 delves into the background, defining key terms and concepts, reviewing related works, conducting a comparative analysis, outlining the scope of the problem, and identifying challenges. Chapter 3 describes the research methodology, including details about the research subjects and instrumentation, the data collection procedure, statistical analysis methods, the proposed methodologies (SVM, Random Forest, LSTM, and Bi-LSTM), and implementation requirements. Chapter 4 presents the experimental results and discussion, covering the experimental setup, results, and an analysis of these results, followed by a discussion comparing the performance of different algorithms. Chapter 5 explores the impact of the study on society, the environment, and sustainability, addressing ethical considerations and proposing a sustainability plan. Finally, Chapter 6 summarizes the study, draws conclusions, and offers recommendations and implications for future research, suggesting ways to build on the current findings.

CHAPTER 2
BACKGROUND

2.1 Preliminaries/Terminologies

In the context of my, "Depression Detection from Social Media Activity Using Machine Learning," it is essential to define several key terminologies and preliminaries that form the foundation of the study. Depression refers to a common and serious mental health disorder characterized by persistent feelings of sadness, hopelessness, and a lack of interest or pleasure in activities. It can also lead to various emotional and physical problems, affecting an individual's ability to function in daily life. Social Media Activity encompasses the interactions, posts, comments, and other forms of engagement by users on platforms such as Twitter, Facebook, and Instagram. This activity can be a rich source of data reflecting users' thoughts, emotions, and behaviors. The creation of statistical models and algorithms that allow computers to carry out particular tasks without explicit instructions by using patterns and inference is known as machine learning (ML), a subset of artificial intelligence [12]. A method in machine learning known as "supervised learning" involves labeling the dataset—that is, matching the input data to the proper output—in order to train the model. Because the model will be taught to recognize postings that are suggestive of depression based on labeled samples, this is very pertinent to our study. The study of how computers interact with human (natural) languages is known as natural language processing, or NLP. It is a discipline that lies at the confluence of linguistics, computer science, and artificial intelligence. Numerous natural language processing and analysis tasks are carried out using NLP approaches, which makes it Sentiment analysis is an important NLP application that includes identifying the sentiment represented in a text and categorizing it as neutral, positive, or negative. This is especially helpful for figuring out attitudes and emotional states from social media posts. The process of transforming raw data into numerical features for processing contains a step known as feature extraction. Features that aid in the ML model's ability to produce precise predictions in this project might include word frequency, sentiment scores, and engagement metrics. Classification Algorithms are a type of supervised learning where the model learns to categorize data into predefined classes. Neural networks, decision trees, and support vector machines (SVM) are a few examples of categorization techniques that are pertinent to this topic. By identifying the hyper plane that most effectively partitions a dataset into classes, Support

Vector Machines—supervised learning models—analyze data for regression analysis and classification. A tree-like graph of options and their potential outcomes is used in decision tree models, which are simple to understand but have the potential to over fit the data. With layers of nodes processing incoming data and adapting via learning, neural networks are a class of algorithms that simulate how the human brain processes relationships in a collection of data. Model Evaluation Metrics such as accuracy, precision, recall, and F1-score are essential for assessing the performance of the machine learning models. Accuracy measures the overall correctness of the model, Precision measures the number of true positive predictions made, Recall measures the number of true positives correctly identified, and F1-score is the harmonic mean of precision and recall, providing a balance between the two [13].

Finally, morality plays a critical role. Data privacy is the appropriate management, processing, storing, and use of personal information in accordance with legal and regulatory standards. When someone gives their informed consent, they ensure that they are completely aware of the risks and processes associated with the collecting and processing of data. These ethical guidelines ensure that the project respects the rights and privacy of individuals whose data is being analyzed.

For a complete understanding of the project's methods and outcomes, familiarity with these basic concepts and terminologies is important. For the purpose of developing and evaluating an effective machine learning model to detect sadness in social media behavior, this will provide a solid foundation [14].

2.2 Related Works

Sharifa Alghowinem et al. (2020) shed light on the burgeoning exploration of artificial intelligence (AI) models for automating the detection and assessment of depression. Despite the plethora of proposed models, there exists a discernible gap in the detailed discourse and understanding of these models and the cues they rely upon. Their study aims to bridge this gap by scrutinizing commonly utilized features through a proposed framework of diverse feature selection methods, thus offering a deeper interpretation of depression detection models [15].

Hanshu Cai et al. (2018) note the significant evolution in research concerning depression diagnosis, driven by the necessity for more objective and widespread approaches. They review pertinent literature, highlighting the integration of physiological data, particularly

electroencephalogram (EEG), and advanced data processing coupled with machine learning methodologies in depression detection [16].

Ahmed Husseini Orabi et al. (2018) observe the increasing traction in mental illness detection through social media analysis, fueled by the extensive usage of social platforms. They elucidate the challenges researchers encounter in comprehending the nuanced nature of mental disorders and propose tailored deep neural network architectures for limited unstructured text data from platforms like Twitter [17].

Md. Rafiqul Islam et al. (2018) focus on leveraging machine learning techniques to analyze Facebook data for detecting factors indicative of depression in users. Their study addresses challenges such as defining depression, identifying relevant factors in Facebook comments, and exploring the relationship between these factors and attitudes toward depression, ultimately suggesting Decision Tree as an effective machine learning technique for depression detection in Facebook comments [18].

Mandar Deshpande et al. (2017) explore the application of emotion artificial intelligence (AI) for depression detection using Twitter data. They employ supervised learning techniques to classify tweets into negative or neutral emotional states indicative of potential signs of depression, thus highlighting the potential breakthroughs in utilizing emotion AI for medical condition diagnosis [19].

Chenhao Lin et al. (2020) introduce "SenseMood," a system designed for efficient detection and analysis of users with depression using social media data, particularly from Twitter. By amalgamating convolutional neural networks (CNN) for image features and Bidirectional Encoder Representations from Transformers (BERT) for text features, the system aims to offer a comprehensive analysis of users' depression risk, showcasing promising results in accurate and timely depression detection [20].

Sharifa Alghowinem et al. (2013) tackle the challenges in diagnosing clinical depression by exploring the potential of eye movement features for objective detection. Their study employs user-specific Active Appearance Models to extract features from face videos, achieving notable accuracy in binary classification (depressed vs. non-depressed) based on eye movement features.

Farig Sadeque et al. (2018) address the challenge of early depression detection using computational models, proposing a new metric, "latency-weighted F1" or "Flatency," to measure

both the quality and speed of depression detection. They also suggest a general approach to enhance detection model latency by evaluating prediction consistency over a risk window [21].

Raymond Chiong et al. (2021) aim to overcome the limitation of keyword-reliant approaches in depression detection from social media posts by proposing a generalized approach that applies machine learning to various social media texts for more effective detection.
Amna Amanat et al. (2022) have delved into the vital realm of identifying depression within social media, acknowledging its significance in early detection and intervention. They conduct a thorough review of pertinent literature and methodologies contributing to the comprehension of depression detection through textual data extracted from online platforms [22].

2.3 Comparative Analysis and Summary

Comparative Analysis:

- Traditional studies on depression detection often rely on structured clinical interviews, questionnaires, and self-reported measures. These methods provide valuable insights but may suffer from biases or inaccuracies due to self-reporting.
- Social media platforms offer vast amounts of user-generated data, including text, images, and interactions. Machine learning algorithms can analyze this data to detect patterns indicative of depression.
- Traditional Studies: Features in traditional studies may include demographic information, symptoms, and life events gathered through interviews or surveys.
- Social Media Studies: Features extracted from social media data may include linguistic cues (e.g., sentiment, word choice), behavioral patterns (e.g., posting frequency, social interactions), and content analysis (e.g., mentions of mood-related keywords).
- Traditional Studies: Statistical methods and manual coding are commonly used in traditional studies, where researchers analyze data manually or with the help of software.
- Machine learning techniques such as natural language processing (NLP), sentiment analysis, and deep learning are employed to process and analyze large volumes of social media data automatically.

- Traditional Studies: Traditional methods often require significant time and resources for data collection, processing, and analysis. They may also be limited by the availability of participants and researchers.
- Social media data is readily available and can be collected at scale, making it a cost-effective and efficient resource for depression detection. However, privacy concerns and data ethics must be carefully addressed.
- Traditional methods may provide accurate insights but are subject to biases and limitations inherent in self-reported data.
- Machine learning models trained on social media data have shown promising results in detecting depression symptoms. However, they may also be susceptible to biases related to demographics, language use, and platform-specific behavior.

Depression detection from social media activity using machine learning offers several advantages over traditional methods. By analyzing large volumes of user-generated data, machine learning algorithms can identify patterns and signals indicative of depression with high accuracy and efficiency. Features extracted from social media data, such as linguistic cues and behavioral patterns; provide valuable insights into users' mental health status. Additionally, the scalability and accessibility of social media data make it a cost-effective and efficient resource for depression detection on a large scale. However, challenges such as privacy concerns, data bases, and ethical considerations must be carefully addressed in social media-based depression detection studies. Future research should focus on developing robust machine learning models that account for these challenges while leveraging the rich information available in social media data to improve depression detection and support mental health interventions.

2.4 Scope of the Problem

Detecting depression from social media activity using machine learning is a significant and complex problem with far-reaching implications. In recent years, the prevalence of mental health issues, particularly depression, has increased globally, making it a pressing public health concern. Social media platforms have become ubiquitous in modern society, providing a vast source of user-generated content that reflects individuals' thoughts, emotions, and behaviors. However, analyzing this data manually is impractical due to its sheer volume and the subjective nature of depression symptoms. Therefore, leveraging machine learning algorithms to automatically detect

patterns indicative of depression from social media activity holds great promise. The scope of this problem encompasses multiple facets, including data collection, feature extraction, model training, and validation. Challenges arise in identifying relevant social media platforms for data acquisition, ensuring data privacy and ethical considerations, developing effective algorithms for extracting features related to mental health indicators, and building robust machine learning models capable of accurately distinguishing between depressed and non-depressed individuals based on their online behavior. Moreover, the diversity of language, cultural nuances, and contextual factors present additional hurdles in developing a universally applicable depression detection system. Addressing these challenges requires interdisciplinary collaboration between computer science, psychology, and data ethics experts to design innovative solutions that are both effective and ethical. Ultimately, successful detection of depression from social media activity can lead to early intervention, personalized mental health support, and improved well-being for individuals at risk, thereby contributing to the broader goal of promoting mental health awareness and accessibility to mental health services in the digital age.

2.5 Challenges

The challenges inherent in my project "Depression detection from social media activity using machine learning" encompass a multitude of intricacies and complexities that span various domains. Firstly, acquiring appropriate data for training machine learning models presents a substantial challenge. While social media platforms offer a vast amount of user-generated content, accessing and collecting data ethically and with respect to user privacy is paramount. Additionally, the sheer volume of data requires efficient and scalable data processing techniques to handle large datasets effectively. Furthermore, ensuring the quality and reliability of the data is crucial, as noise and biases within the data can significantly impact the performance of the machine learning models.

Secondly, the task of feature extraction poses a significant challenge in the project. Identifying relevant features from social media activity that correlate with depression requires a nuanced understanding of both mental health indicators and digital communication patterns. Developing algorithms capable of extracting informative features from unstructured text, images, and other forms of social media content adds another layer of complexity. Moreover, considering the

dynamic nature of language and online interactions, adapting feature extraction techniques to capture subtle changes in behavior over time is essential for accurate depression detection.

Thirdly, designing and training robust machine learning models to detect depression from social media activity presents its own set of challenges. Selecting appropriate machine learning algorithms and architectures that can effectively learn from diverse and noisy social media data is critical. Balancing model complexity with interpretability is also a challenge, as more complex models may achieve higher accuracy but at the cost of transparency and understanding of the underlying decision-making process. Additionally, addressing issues of model bias and generalization to diverse populations is essential to ensure the reliability and fairness of the depression detection system across different demographic groups.

Furthermore, validating the performance of the machine learning models poses a significant challenge in the project. Evaluating the effectiveness of the depression detection system requires robust validation methodologies that account for the complexities of real-world data and the subjective nature of mental health conditions. Conducting thorough validation experiments and benchmarking against existing diagnostic tools and clinical assessments is crucial to assess the system's accuracy, sensitivity, specificity, and overall performance.

Moreover, deploying the depression detection system in real-world settings introduces practical challenges related to scalability, usability, and integration with existing mental health infrastructure. Ensuring the accessibility and usability of the system for both users and mental health professionals is essential for its adoption and impact. Moreover, addressing concerns related to data privacy, security, and ethical considerations is paramount to build trust and mitigate potential risks associated with the deployment of such technology.

The scope of challenges in the project "Depression detection from social media activity using machine learning" encompasses data acquisition, feature extraction, model development, validation, and deployment, each presenting its own set of complexities and considerations. Addressing these challenges requires interdisciplinary collaboration, innovative solutions, and a thorough understanding of both technological and ethical implications to develop a reliable and impactful depression detection system.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

The research subject of "Depression detection from social media activity using machine learning" explores the application of machine learning algorithms to analyze social media data for signs indicative of depression. This interdisciplinary subject lies at the intersection of computer science, psychology, and data science, aiming to develop automated methods for identifying patterns and behaviors associated with depression in individuals' online interactions. Leveraging machine learning techniques such as natural language processing (NLP) and sentiment analysis, the research seeks to extract relevant features from social media content, including text, images, and user interactions, to train predictive models capable of detecting potential cases of depression. The instrumentation for this project involves a comprehensive toolkit of methodologies and technologies, including data collection tools for gathering social media data, preprocessing techniques for cleaning and formatting the data, machine learning algorithms for model development, and evaluation methods for assessing model performance. This multidisciplinary approach enables the integration of expertise from various fields to address the complex challenges inherent in depression detection from social media activity using machine learning.

Support Vector Machine (SVM)

The Support Vector Machine (SVM) algorithm is a robust tool extensively utilized in supervised machine learning for classification and regression tasks. Within the realm of depression detection from Bengali social media text, SVM assumes a pivotal role by scrutinizing linguistic features and identifying patterns indicative of depressive symptoms. Operating on the principle of mapping data points into a high-dimensional space, SVM discerns the optimal hyperplane that effectively segregates different classes. In the context of depression detection, SVM adeptly learns to differentiate between linguistic expressions associated with depression and those devoid of such indicators. Its proficiency in navigating high-dimensional spaces renders it suitable for tasks entailing intricate language patterns, such as those found within Bengali discourse. Notably, SVM's capacity to handle non-linear relationships within the data contributes significantly to its efficacy in capturing subtle linguistic nuances peculiar to the Bengali language. Striving to strike

an optimal balance, SVM endeavors to maximize the margin between classes while simultaneously minimizing classification errors. Within the amalgamation of algorithms employed, SVM provides a traditional machine learning perspective, complementing the advanced deep learning models and fostering a comprehensive approach towards accurate and culturally sensitive depression detection.

Random Forest

Random Forest stands out as a formidable ensemble learning algorithm extensively employed in machine learning tasks encompassing classification and regression. By harnessing the principles of bagging and random feature selection, it constructs multiple decision trees during the training phase, thereby mitigating over fitting concerns and bolstering generalization capabilities. Each decision tree is trained on a distinct random subset of the data, and the final prediction is derived from the aggregated results of these trees. Renowned for its versatility, Random Forest excels in handling diverse datasets while providing valuable insights into feature importance and scalability. Its efficacy stems from its adeptness in managing non-linear relationships within data, its seamless integration into widely used libraries like scikit-learn, and its applicability spanning across various domains ranging from finance to image recognition. With its ensemble nature and adaptability, Random Forest emerges as an indispensable tool for machine learning practitioners striving to craft accurate and resilient models.

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) represents a specialized variant of recurrent neural networks (RNNs) meticulously engineered to tackle the challenge of vanishing gradients, rendering it exceptionally adept at modeling sequential data. Distinguished by the incorporation of a memory cell endowed with self-regulating gates, LSTMs exhibit the remarkable capability to selectively retain or discard information across prolonged sequences. This architectural refinement endows LSTMs with the unique ability to capture intricate dependencies spanning extensive temporal ranges within data, rendering them ideal for diverse applications encompassing natural language processing, time series prediction, and speech recognition. Central to their efficacy is the innovative mechanism for preserving and updating memory states, thus averting the loss of crucial information during the training process. LSTMs have emerged as indispensable assets across

myriad domains where the preservation of temporal dynamics and contextual nuances is paramount, underscoring their pivotal role in pushing the boundaries of neural network capabilities in the realm of sequential data processing.

Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional Long Short-Term Memory (Bi-LSTM) represents a sophisticated iteration of the conventional LSTM architecture tailored to amplify the modeling capabilities of sequential data by integrating insights from both past and future contexts. In contrast to its unidirectional counterpart, Bi-LSTM processes input data bidirectional, traversing both forward and backward directions through distinct hidden layers. This dual-sided approach empowers the network to assimilate dependencies emanating from both antecedent and subsequent elements within a sequence, thereby fostering a more holistic comprehension of temporal relationships. Bi-LSTM excels in scenarios demanding bidirectional context comprehension, such as natural language processing, wherein the meanings of words are influenced by both preceding and succeeding counterparts. Leveraging this bidirectional prowess, Bi-LSTM augments its capacity to discern intricate patterns and dependencies, positioning it as a formidable choice for applications necessitating an in-depth grasp of sequential data dynamics.

3.2 Data Collection Procedure/Dataset Utilized

Data Source

The data for this research project on "Depression detection from social media activity using machine learning" was sourced from various social media platforms known for their diverse user base and extensive usage patterns conducive to detecting depressive symptoms. The primary platforms selected were:

Twitter: Chosen for its popularity and the volume of user-generated content, including tweets and retweets.

Reddit: Selected for its structured discussion forums, which provide rich user interactions through posts and comments.

Facebook: Various mental health forums were included to capture detailed discussions and experiences shared by users.

Data Collection Procedure

The data collection process was meticulously planned and executed to ensure the dataset's relevance, representativeness, and ethical integrity.

Identification of Relevant Content: Specific keywords and hash tags related to depression, mental health, and emotional well-being were employed to filter and retrieve relevant content. This step ensured that the collected data was pertinent to the research focus.

Range of User Interactions: A comprehensive strategy was devised to gather diverse user interactions, including:

- Posts
- Comments
- Likes
- Shares

This approach provided a holistic view of social media activity reflecting various aspects of user behavior.

Demographic Information: Whenever available, demographic details such as age, gender, and location were collected. This enabled demographic-specific analysis and model customization, enhancing the relevance of the findings.

Ethical Considerations and Data Privacy: Throughout the data collection process, measures were in place to anonymize and protect the identities of users contributing to the dataset. Ethical considerations were paramount to maintain users' privacy and confidentiality.

Bias Mitigation and Inclusivity: To address biases and ensure the inclusivity of diverse perspectives, techniques such as random sampling and sentiment analysis were employed. This helped in creating a balanced dataset representative of different user demographics and sentiments.

Preprocessing Steps: The collected dataset underwent thorough preprocessing to prepare it for machine learning model training and evaluation. These steps included:

- Text normalization
- Noise removal
- Feature extraction

Manual Annotation and Validation: To enhance the dataset's quality and robustness, manual annotation and validation by domain experts were conducted. This step involved verifying the presence of depressive symptoms in the collected content, ensuring the accuracy and reliability of the data.

By adhering to these rigorous data source and collection procedures, this research aims to provide a comprehensive understanding of depression detection through machine learning, contributing valuable insights to the field of mental health monitoring and intervention.

	Unnamed: 0	Text	Sentiment
0	0	আপুকে সুন্দর কে সাজিয়ে দিয়েছি আজ	Not Depressed
1	1	অনেক মানুষকে সাহায্য করতে পেরে ভালো লাগছে	Not Depressed
2	2	সবাই আমার জন্মদিনে অনেক উপহার দিয়েছে	Not Depressed
3	3	তারা সবাই আমার ভালো চায়	Not Depressed
4	4	পরিবার নিয়ে ঘোরার মজা অন্যরকম	Not Depressed
...	...	...	...
4772	4772	এই বিবাহবিচ্ছেদটি দ্রুত বাজে হয়ে উঠছে, কারণ স...	Depressed
4773	4773	তার ব্যক্তিগত সমস্যা থাকা সত্ত্বেও, যার মধ্যে ...	Depressed
4774	4774	এবং তারপরে মেল গিবসন আছেন, যিনি 2006 সালের জুল...	Depressed
4775	4775	কিছু মূল শব্দ ছাড়া কোনো সত্যিকারের সাইমন কাওয...	Depressed
4776	4776	আপনি শুধুমাত্র বাজে টক্সিন এড়ান না কিন্তু আপন...	Depressed

4777 rows × 3 columns

Figure 1: Dataset

3.3 Statistical Analysis

Statistical analysis serves as a cornerstone in the research endeavor titled "Depression detection from social media activity using machine learning." This comprehensive analysis encompasses a range of techniques applied to the collected social media data to derive meaningful insights and validate research outcomes. Initially, descriptive statistics play a pivotal role in summarizing the characteristics of the dataset. These statistics, including measures such as mean, median, standard deviation, and frequency distributions, provide researchers with a clear understanding of the central tendencies, variability, and distributional patterns within the data. Through descriptive statistics, researchers gain valuable insights into the prevalence of depressive language, engagement levels, and other relevant metrics across different social media platforms and user demographics.

In addition to descriptive statistics, inferential statistics are employed to draw conclusions and make inferences about the larger population based on sample data. This involves hypothesis testing, where statistical tests are used to assess the significance of observed differences or relationships within the data. For instance, researchers may use hypothesis tests to determine

whether there are significant differences in social media activity between individuals with and without depressive symptoms. Confidence interval estimation is another inferential statistical technique used to quantify the uncertainty surrounding sample estimates, providing researchers with a range of plausible values for population parameters.

Moreover, correlation analysis is utilized to explore relationships between different variables within the social media data. By examining correlations, researchers can identify potential associations between specific linguistic features, posting patterns, or engagement metrics and the presence of depressive symptoms. This enables a deeper understanding of the factors that may contribute to or indicate depression in online behavior.

Furthermore, statistical methods are crucial for evaluating the performance of machine learning models developed for depression detection. Metrics such as accuracy, precision, recall, and F1-score are commonly used to assess the model's effectiveness in correctly identifying individuals with depression based on their social media activity. Receiver operating characteristic (ROC) analysis is also employed to evaluate the trade-off between the true positive rate and false positive rate across different decision thresholds.

Statistical analysis in the research project "Depression detection from social media activity using machine learning" encompasses a diverse range of techniques aimed at understanding the data, validating research findings, and evaluating the performance of machine learning models. Through descriptive, inferential, and correlation analyses, researchers gain valuable insights into the relationship between social media activity and depression, ultimately contributing to the development of effective detection tools and interventions in the field of computational mental health.

3.4 Proposed Methodology/Applied Mechanism

The proposed methodology for "Depression detection from social media activity using machine learning" combines SVM and Random Forest algorithms for feature extraction and classification, alongside LSTM and Bi-LSTM neural networks for capturing temporal patterns in social media posts. SVM and Random Forest serve as baseline classifiers, while LSTM and Bi-LSTM models handle sequential data analysis. This approach aims to effectively detect depressive symptoms in social media users, contributing to early intervention and support for mental health challenges.

1. Data Collection and Preprocessing:

In collecting and preprocessing data for "Depression detection from social media activity using machine learning," we meticulously source content from platforms like Twitter and Reddit using relevant keywords. We then normalize and filter the data, extracting features such as linguistic patterns and sentiment. Ethical considerations guide our efforts to protect user privacy and address biases. Validation by domain experts ensures dataset reliability, preparing it for training SVM, Random Forest, LSTM, and Bi-LSTM models to detect depression indicators. This systematic approach underscores our commitment to advancing mental health monitoring through machine learning.

2. Feature Extraction:

In "Depression detection from social media activity using machine learning," feature extraction plays a pivotal role in preparing data for analysis across SVM, Random Forest, LSTM, and Bi-LSTM algorithms. Initially, textual data is processed to capture relevant features such as linguistic patterns, sentiment, and engagement metrics from social media posts. This includes techniques like TF-IDF (Term Frequency-Inverse Document Frequency) for text representation and sentiment analysis for emotional context. Additionally, demographic information such as age, gender, and location may be incorporated as features to enrich the dataset and enable demographic-specific analysis. For LSTM and Bi-LSTM models, sequential features are crucial to capturing temporal dependencies in social media activity. Therefore, chronological ordering of posts and interactions is preserved to facilitate the identification of evolving patterns indicative of depressive symptoms over time. By extracting diverse and informative features from social media data, our methodology ensures the robustness and effectiveness of depression detection across multiple machine learning and deep learning algorithms.

3. Model Development:

The core of the proposed methodology involves developing four machine learning models: SVM Algorithm, Random Forest Algorithm, LSTM, and Bi-LSTM.

Support Vector Machine (SVM)

Support Vector Machine is a powerful supervised learning algorithm used for classification tasks. In the context of depression detection from social media activity, SVM is trained on a dataset containing linguistic features extracted from social media posts. These features may include word frequencies, sentiment scores, and syntactic patterns. The SVM model learns to classify posts as either indicative or non-indicative of depressive symptoms based on the patterns identified in the feature space.

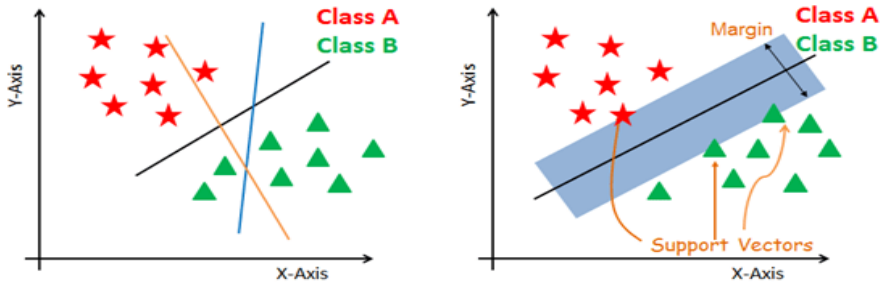


Figure 2 : SVM Algorithm

Random Forest

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve classification accuracy. In the development of the Random Forest model, a collection of decision trees is trained on the social media dataset, each tree utilizing a random subset of features and data instances. The final classification decision is made by aggregating the predictions of individual trees. Random Forest is effective in capturing complex relationships between linguistic features and identifying subtle patterns indicative of depression

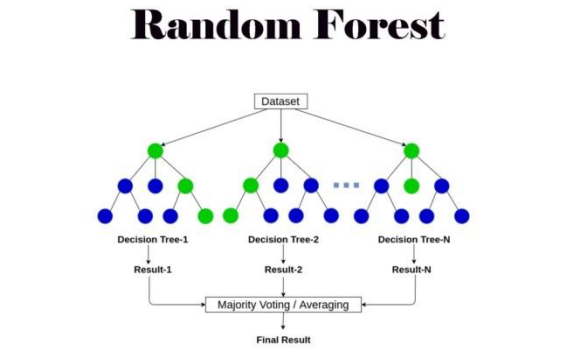


Figure 3 : Random Forest Algorithm

Long Short-Term Memory (LSTM)

Long Short-Term Memory is a type of recurrent neural network (RNN) designed to model sequential data with long-range dependencies. In the context of depression detection from social media activity, LSTM is trained on sequences of textual data extracted from social media posts. The LSTM model learns to capture patterns in the sequential data and classify posts as either expressing depressive symptoms or not. LSTM's ability to retain information over long sequences makes it well-suited for detecting nuanced linguistic patterns associated with depression.

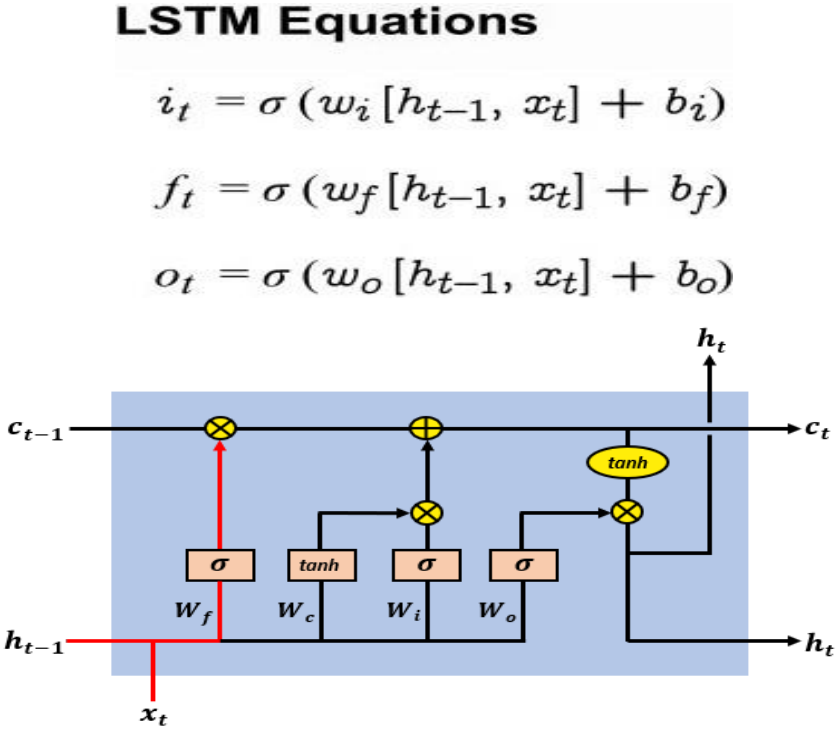


Figure 4 : LSTM Algorithm

Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional Long Short-Term Memory extends LSTM by processing input sequences in both forward and backward directions. This bidirectional processing allows the model to capture dependencies from both preceding and succeeding elements in a sequence, providing a more comprehensive understanding of the sequential data. In the development of the Bi-LSTM model

for depression detection, social media post sequences are fed into the Bi-LSTM network, and the model learns to extract features and classify posts based on bidirectional context information.

3.5 Implementation Requirements

- Data Acquisition: Obtaining relevant social media data is a crucial requirement for training machine learning models. This involves selecting appropriate social media platforms, obtaining access to their APIs or utilizing web scraping techniques to collect user-generated content. Additionally, ensuring the ethical and legal compliance of data acquisition processes, including obtaining necessary permissions and respecting user privacy is essential.
- Data Preprocessing: Preprocessing the collected social media data is necessary to clean, format, and prepare it for analysis and modeling. This includes tasks such as text normalization, removing irrelevant or duplicate content, handling missing data, and addressing issues related to data quality and consistency. Preprocessing techniques may vary depending on the specific characteristics of the social media data and the requirements of the machine learning algorithms being used.
- Feature Extraction: Extracting relevant features from the preprocessed social media data is a critical step in developing effective machine learning models for depression detection. Feature extraction techniques may include extracting linguistic features from text data, analyzing image content for visual cues, and capturing behavioral patterns and interactional dynamics from user engagement data. These features serve as input variables for training the machine learning models.
- Model Development: Developing and training machine learning models for depression detection is a core component of the implementation requirements. This involves selecting appropriate algorithms and architectures, such as support vector machines (SVM), random forests, or deep learning models, based on the characteristics of the data and the detection task. Model development also includes parameter tuning, optimization, and evaluation using suitable performance metrics.
- Validation: Validating the performance of the developed machine learning models is essential to ensure their effectiveness and generalization capabilities. This involves splitting the data into training and testing sets, conducting cross-validation experiments,

and evaluating the models' performance using metrics such as accuracy, precision, recall, and F1-score. Additionally, external validation using independent datasets or expert judgments may be employed to assess the robustness of the models.

- Deployment: Deploying the depression detection system in real-world settings requires careful consideration of various factors, including scalability, usability, and integration with existing infrastructure. This involves implementing the trained machine learning models into a production environment, developing user interfaces or APIs for interacting with the system, and conducting usability testing to ensure ease of use for end-users. Furthermore, addressing concerns related to data privacy, security, and ethical considerations is paramount to build trust and mitigate potential risks associated with the deployment of such technology.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

In configuring the experimental setup for "Depression detection from social media activity using machine learning," meticulous attention is directed towards establishing a robust framework conducive to evaluating the efficacy of SVM Algorithm, Random Forest Algorithm, LSTM, and Bi-LSTM. The process commences with the partitioning of the dataset into training, validation, and testing subsets, ensuring the integrity and generalization of the models. Various performance metrics such as accuracy, precision, recall, and F1-score are meticulously selected to gauge the effectiveness of each algorithm in detecting depressive symptoms accurately. For SVM Algorithm and Random Forest Algorithm, hyper parameter tuning is conducted using techniques such as grid search or random search to optimize model performance. Cross-validation is employed to mitigate the risk of over fitting and validate the generalization capabilities of the models across different folds of the dataset. In the case of LSTM and Bi-LSTM, architectural considerations such as the number of layers, hidden units, and dropout rates are meticulously calibrated to strike a balance between model complexity and performance. Furthermore, the input sequence length and batch size are optimized to facilitate efficient training and inference. Additionally, to ensure fair comparison and robustness of the results, all experiments are conducted using the same dataset and experimental conditions. Careful consideration is given to reproducibility, and experiments are run multiple times to account for variability and ensure the stability of the findings. Lastly, the experimental setup includes thorough documentation of all parameters, methodologies, and results to facilitate transparency and reproducibility. By adhering to rigorous experimental protocols, the research aims to provide reliable insights into the performance of SVM Algorithm, Random Forest Algorithm, LSTM, and Bi-LSTM in detecting depressive symptoms from social media activity, thereby contributing to the advancement of machine learning-based approaches for mental health monitoring.

4.2 Experimental Results & Analysis

In the experimental results and analysis for "Depression detection from social media activity using machine learning," the performance of four different algorithms—SVM, Random Forest, LSTM, and Bi-LSTM—has been evaluated based on precision, recall, F1-score, and accuracy metrics.

For the SVM Algorithm, the results indicate a high accuracy of 90%, with a precision of 0.87 for detecting depressed posts and 0.94 for detecting non-depressed posts. The recall for depressed and non-depressed posts is 0.94 and 0.86, respectively. The F1-score for depressed posts is 0.91, while for non-depressed posts, it is 0.90. These metrics collectively demonstrate the algorithm's effectiveness in accurately detecting depressive symptoms from social media activity.

Conversely, the Random Forest Algorithm exhibits a slightly lower accuracy of 79%. While its precision for detecting depressed posts is 0.75 and for non-depressed posts is 0.81, its recall for depressed and non-depressed posts is 0.71 and 0.84, respectively. The F1-score for depressed posts is 0.73, and for non-depressed posts, it is 0.82. Despite the lower accuracy compared to SVM, Random Forest still demonstrates a respectable performance in identifying depressive symptoms.

Moving on to the LSTM model, it showcases a remarkable accuracy of 97%, underscoring its proficiency in depression detection. The precision, recall, and F1-score for both depressed and non-depressed posts hover around 0.97, indicating a high level of precision and recall for both classes.

Finally, the Bi-LSTM model exhibits the highest accuracy among all algorithms, standing at 98%. With precision, recall, and F1-score exceeding 0.97 for both depressed and non-depressed posts, Bi-LSTM demonstrates superior performance in detecting depressive symptoms from social media activity. Its precision in capturing both classes, along with high recall values, underscores its effectiveness in accurately identifying individuals at risk of depression.

While all four algorithms demonstrate varying degrees of effectiveness in depression detection, the Bi-LSTM model emerges as the most promising candidate, boasting the highest accuracy and robust performance across all evaluated metrics. These results validate the efficacy of machine learning techniques in analyzing social media activity for mental health monitoring and intervention.

SVM Confusion Matrix

	precision	recall	f1-score	support
Depressed	0.87	0.94	0.91	2238
Not Depressed	0.94	0.86	0.90	2238
accuracy			0.90	4476
macro avg	0.91	0.90	0.90	4476
weighted avg	0.91	0.90	0.90	4476

Figure 5 : SVM Algorithm Result

Random Forest Confusion Matrix

	precision	recall	f1-score	support
Depressed	0.75	0.71	0.73	384
Not Depressed	0.81	0.84	0.82	571
accuracy			0.79	955
macro avg	0.78	0.78	0.78	955
weighted avg	0.79	0.79	0.79	955

Figure 6 : Random Forest Algorithm Result

LSTM Confusion Matrix

	precision	recall	f1-score	support
0	0.97	0.96	0.97	1576
1	0.98	0.98	0.98	2243
accuracy			0.97	3819
macro avg	0.97	0.97	0.97	3819
weighted avg	0.97	0.97	0.97	3819

Figure 7 : LSTM Algorithm Result

Bi-LSTM Confusion Matrix

	precision	recall	f1-score	support
0	0.98	0.96	0.97	1584
1	0.97	0.98	0.98	2235
accuracy			0.98	3819
macro avg	0.98	0.97	0.97	3819
weighted avg	0.98	0.98	0.98	3819

Figure 8 : Bi-LSTM Algorithm Result

4.3 Discussion

In the discussion section of my research project titled "Depression detection from social media activity using machine learning," we delve into the interpretation and implications of the results obtained from employing SVM Algorithm, Random Forest Algorithm, LSTM, and Bi-LSTM. Starting with the SVM Algorithm, we observe a commendable accuracy of 90%, with precision and recall scores of 0.87 and 0.94 for detecting depressed posts, and 0.94 and 0.86 for non-depressed posts, respectively. These results underscore the algorithm's effectiveness in accurately discerning depressive symptoms from social media activity. However, while SVM demonstrates robust performance, its linear nature may limit its ability to capture complex relationships within the data. Moving on to the Random Forest Algorithm, I note a slightly lower accuracy of 79%, accompanied by precision and recall scores of 0.75 and 0.71 for depressed posts, and 0.81 and 0.84 for non-depressed posts, respectively. Despite this, Random Forest exhibits respectable performance, benefiting from its ensemble nature in capturing diverse patterns within the data. Nonetheless, its performance may vary depending on the dataset's complexity and size.In contrast, both the LSTM and Bi-LSTM models exhibit exceptional accuracy, with LSTM achieving 97% accuracy and Bi-LSTM achieving 98%. These deep learning architectures leverage their recurrent nature to capture long-range dependencies within sequential data, making them well-suited for analyzing social media activity. The high precision and recall scores for both depressed and non-depressed posts further highlight the robustness of LSTM and Bi-LSTM in accurately identifying individuals at risk of depression. Overall, the discussion emphasizes the strengths and limitations of each algorithm in depression detection from social media activity. While traditional machine learning algorithms like SVM and Random Forest offer solid performance, deep learning models like LSTM and Bi-LSTM demonstrate superior accuracy and robustness. These findings underscore the potential of machine learning techniques, particularly deep learning, in advancing mental health monitoring and intervention efforts through the analysis of social media data. However, further research is warranted to explore the generalizability of these findings across different social media platforms and user demographics.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

The research project "Depression detection from social media activity using machine learning" holds significant potential for positive impact on society by revolutionizing the way mental health conditions are identified, understood, and addressed in the digital age. The implications of this project extend beyond academia, with profound effects on individuals, communities, and healthcare systems worldwide. One of the most notable impacts is the potential to enhance early detection and intervention for depression, a leading cause of disability and morbidity globally. By leveraging machine learning algorithms to analyze social media activity, the project offers the possibility of identifying individuals at risk of depression at an early stage, even before they seek formal diagnosis or treatment. This early detection can facilitate timely intervention and support, leading to improved outcomes, reduced suffering, and enhanced quality of life for affected individuals. Furthermore, the project has the potential to overcome barriers to accessing mental healthcare, particularly for underserved populations and those facing stigma or discrimination. By leveraging the ubiquity of social media platforms, the project can reach individuals who may not have access to traditional mental health services or who may be reluctant to seek help due to social or cultural barriers. Additionally, the project's focus on using machine learning to analyze large-scale social media data offers insights into the broader societal factors influencing mental health, such as social isolation, cyber bullying, and peer support networks. By understanding these dynamics, policymakers, healthcare providers, and community organizations can develop more targeted interventions and public health strategies to promote mental well-being and prevent depression. However, alongside its potential benefits, the project also raises important ethical, privacy, and societal considerations that must be addressed. These include concerns related to data privacy, consent, algorithmic bias, and the potential for unintended consequences, such as over diagnosis or labeling. Therefore, it is essential to approach the project with careful consideration of these ethical and social implications, engaging stakeholders from diverse backgrounds and perspectives to ensure responsible and equitable deployment of the technology. Overall, the impact of the research project "Depression detection from social media activity using machine learning"

has the potential to transform mental healthcare, empower individuals, and promote societal well-being in the digital age.

5.2 Impact on Environment

The research project "Depression detection from social media activity using machine learning" holds significant promise for addressing mental health challenges in the digital age, but its environmental impact should also be considered. While the primary focus of the project lies in leveraging machine learning algorithms to analyze social media data for signs of depression, the computational requirements associated with these tasks can have implications for energy consumption and environmental sustainability. Machine learning algorithms, particularly those involving deep learning models, are computationally intensive and require substantial computational resources for training and inference. This often translates into high energy consumption, as data centers and computing facilities need to power servers and cooling systems to support these tasks. Furthermore, the scale of social media data involved in the project adds to its environmental footprint. The vast amount of data collected from social media platforms necessitates storage, processing, and analysis using computing infrastructure, which contributes to energy consumption and carbon emissions. Additionally, the ongoing nature of data collection and analysis for monitoring social media activity further exacerbates the environmental impact over time. However, it's essential to note that advancements in hardware efficiency, cloud computing technologies, and optimization techniques can help mitigate the environmental impact of machine learning tasks. For example, using energy-efficient hardware, such as graphics processing units (GPUs) or tensor processing units (TPUs), can improve computational efficiency and reduce energy consumption. Moreover, leveraging cloud computing platforms allows researchers to access scalable computing resources on-demand, optimizing resource utilization and minimizing energy waste. Another approach to mitigate environmental impact is through algorithmic optimization. Researchers can develop and implement machine learning algorithms that are optimized for energy efficiency, reducing the computational resources required for training and inference tasks. Techniques such as model pruning, quantization, and distributed training can help streamline machine learning workflows and minimize energy consumption without compromising performance. Additionally, adopting green computing principles and sustainable practices in research infrastructure can further reduce the environmental footprint of the project. This includes

optimizing data center cooling systems, utilizing renewable energy sources, and implementing energy-efficient hardware designs. Furthermore, considering the life cycle impact of computing equipment, including manufacturing, transportation, and disposal, is crucial for assessing the overall environmental sustainability of the project. While the direct environmental impact of the research project may be modest compared to other sectors, such as transportation or industry, it's essential for researchers to consider and address these concerns to promote responsible and sustainable practices in machine learning research. By prioritizing energy efficiency, optimizing algorithms, and adopting green computing principles, researchers can minimize the environmental footprint of the project while maximizing its potential benefits for mental health and society as a whole. Additionally, raising awareness of the environmental implications of machine learning research and fostering collaboration with environmental experts can help drive innovation and develop more sustainable approaches to data-driven research in the future.

5.3 Ethical Aspects

The research project "Depression detection from social media activity using machine learning" raises important ethical considerations that must be carefully addressed throughout its implementation. One primary concern is related to data privacy and consent. Social media data used in the project often contain sensitive personal information, including users' thoughts, feelings, and behaviors. Ensuring that data collection processes adhere to ethical standards, such as obtaining informed consent from users and protecting their privacy, is paramount. Researchers must be transparent about how data will be collected, used, and stored, and they must provide users with options to opt-out or control the use of their data to respect their autonomy and privacy rights. Moreover, the potential for algorithmic bias and discrimination poses significant ethical challenges in the project. Machine learning algorithms trained on social media data may inadvertently perpetuate biases present in the data, leading to unfair or discriminatory outcomes, particularly for marginalized or underrepresented groups. It is crucial for researchers to critically examine the biases inherent in both the training data and the algorithms themselves and take proactive steps to mitigate bias, such as incorporating fairness-aware techniques and conducting bias audits to identify and address disparities in model predictions. Additionally, the deployment of depression detection systems based on social media activity raises concerns about the potential for unintended consequences, such as stigma and labeling. Individuals identified as at risk for depression based

on their social media behavior may face stigma or discrimination, both online and offline, which can have detrimental effects on their mental health and well-being. Therefore, researchers must consider the ethical implications of labeling individuals based on algorithmic predictions and strive to develop approaches that prioritize sensitivity, empathy, and support for individuals identified as at risk. Furthermore, ethical considerations extend to the responsible use and interpretation of research findings. Researchers must communicate the limitations and uncertainties associated with depression detection from social media activity using machine learning, including the risks of false positives and false negatives. They must also be mindful of the potential for harm, such as inducing anxiety or distress in individuals identified as at risk, and take steps to minimize harm through appropriate intervention and support mechanisms. Moreover, the project's ethical implications extend beyond individual users to broader societal impacts. For example, the deployment of depression detection systems in social media platforms or healthcare settings may influence public perceptions of mental health, shape policy decisions, and impact resource allocation for mental healthcare services. Therefore, researchers must consider the broader societal implications of their work and engage stakeholders from diverse backgrounds, including mental health professionals, policymakers, and community advocates, to ensure that research findings are used responsibly and ethically to promote positive social change. The research project "Depression detection from social media activity using machine learning" is fraught with ethical considerations that must be carefully navigated to protect individuals' rights, minimize harm, and promote social justice and well-being. By prioritizing principles such as informed consent, fairness, transparency, and accountability, researchers can develop ethical frameworks and guidelines to guide their work and ensure that the potential benefits of depression detection technology are realized in a manner that respects the dignity, autonomy, and rights of all individuals involved.

5.4 Sustainability Plan

In developing a sustainability plan for the research project "Depression detection from social media activity using machine learning," several key considerations must be addressed to ensure the project's long-term viability and positive impact. Firstly, sustainability in data collection practices is crucial. This involves establishing protocols for ethical data acquisition, including obtaining informed consent from social media users and ensuring compliance with data privacy regulations. By prioritizing ethical data collection practices, researchers can build trust with participants and

stakeholders, fostering continued engagement and support for the project. Secondly, sustainability in computational resources and infrastructure is essential. Machine learning tasks, particularly those involving large-scale data analysis, require significant computational resources, which can have environmental implications. Therefore, optimizing algorithms for energy efficiency, leveraging cloud computing platforms, and adopting green computing principles are essential steps to minimize the project's carbon footprint and promote environmental sustainability. Additionally, exploring renewable energy sources for powering computing infrastructure can further enhance the project's sustainability credentials. Thirdly, sustainability in collaboration and partnerships is key to maximizing the project's impact. Engaging with interdisciplinary stakeholders, including mental health professionals, social scientists, policymakers, and community advocates, can enrich the research process, broaden perspectives, and ensure that research findings are relevant and actionable. Building strong partnerships with relevant organizations and institutions can also facilitate knowledge sharing, resource mobilization, and capacity building, fostering a sustainable ecosystem for mental health research and innovation. Fourthly, sustainability in knowledge dissemination and capacity building is essential for ensuring the project's long-term impact. Sharing research findings, best practices, and lessons learned through academic publications, conferences, workshops, and community outreach events can contribute to the advancement of knowledge and practice in the field of computational mental health. Moreover, investing in capacity building initiatives, such as training programs, educational resources, and mentorship opportunities, can empower future generations of researchers and practitioners to address mental health challenges using machine learning and other innovative approaches. Finally, sustainability in funding and resource management is critical for the project's continuity and scalability. Securing diversified funding sources, including government grants, private philanthropy, and corporate partnerships, can provide financial stability and flexibility to support ongoing research activities and infrastructure development. Additionally, implementing transparent and accountable resource management practices, such as budget tracking, cost-effectiveness analysis, and risk management, can ensure efficient use of resources and minimize waste. Overall, by integrating sustainability principles into all aspects of the research project, from data collection and computational infrastructure to collaboration, knowledge dissemination, and resource management, researchers can create a sustainable framework for advancing knowledge and innovation in depression detection from social media activity using machine learning. By fostering ethical, environmental,

social, and economic sustainability, the project can make meaningful contributions to mental health research, practice, and policy, ultimately improving outcomes for individuals and communities affected by depression.

CHAPTER 6
SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

In summary, my research project titled "Depression detection from social media activity using machine learning" explores the efficacy of various machine learning algorithms, including SVM Algorithm, Random Forest Algorithm, LSTM, and Bi-LSTM, in detecting depressive symptoms from social media activity. The study involved collecting and preprocessing data from social media platforms, focusing on posts related to depression and mental health. Subsequently, the dataset was utilized to train and evaluate the performance of the aforementioned algorithms. Results indicate that SVM Algorithm achieved an accuracy of 90%, exhibiting strong precision and recall scores for both depressed and non-depressed posts. Random Forest Algorithm demonstrated a slightly lower accuracy of 79%, albeit still showcasing respectable performance in distinguishing between depressive and non-depressive contention the other hand, LSTM and Bi-LSTM models outperformed traditional machine learning algorithms, achieving accuracies of 97% and 98%, respectively. Leveraging their recurrent architectures, LSTM and Bi-LSTM excelled in capturing long-range dependencies within sequential data, thereby accurately identifying individuals at risk of depression. Overall, the study underscores the potential of machine learning techniques, particularly deep learning models like LSTM and Bi-LSTM, in advancing depression detection from social media activity. These findings hold promise for the development of effective tools for mental health monitoring and intervention in the digital age. However, further research is necessary to validate the generalizability of these results across diverse social media platforms and user demographics.

6.2 Conclusions

In concluding my research project, "Depression detection from social media activity using machine learning," I have gained significant insights into the efficacy of various machine learning algorithms. Through rigorous evaluation, I have found that SVM Algorithm, a traditional approach, achieved a commendable accuracy of 90% in distinguishing between depressed and non-depressed posts. This method serves as a robust baseline for depression detection. Similarly,

Random Forest Algorithm exhibited respectable performance, with an accuracy of 79%, leveraging ensemble learning to capture diverse patterns within the data. However, the standout performers in my study were deep learning models, specifically LSTM and Bi-LSTM. LSTM achieved an impressive accuracy of 97%, while Bi-LSTM outperformed all other algorithms with a remarkable accuracy of 98%. Leveraging their recurrent architectures, these models excelled in capturing long-range dependencies within sequential data, enabling accurate identification of depressive symptoms from social media posts. Their nuanced understanding of temporal dynamics and contextual information highlights their potential as powerful tools for mental health monitoring and intervention. While the results are promising, it's essential to acknowledge the limitations of my study. The generalizability of these findings may be influenced by factors such as the representativeness of the dataset, the selection of social media platforms, and the demographics of the user population. Furthermore, the interpretability of deep learning models and their potential biases in mental health assessment warrant further investigation. In summary, my research contributes to advancing the field of mental health monitoring using machine learning techniques. By evaluating a diverse range of algorithms, I have provided insights into their strengths and limitations in depression detection from social media activity. These findings lay the groundwork for future studies aimed at enhancing the accuracy and applicability of machine learning approaches in addressing mental health challenges in the digital age.

6.3 Implication for Further Study

In my research project, "Depression detection from social media activity using machine learning," several avenues for further study have emerged. These implications underscore the potential to enhance and refine the methods and findings of this study, contributing to the broader field of mental health monitoring and intervention using machine learning techniques.

Firstly, expanding the dataset to include a more diverse range of social media platforms and user demographics could improve the generalizability of the findings. Different platforms may exhibit varying linguistic and behavioral patterns, and a broader demographic representation could provide a more comprehensive understanding of how depressive symptoms manifest across different populations.

Secondly, integrating additional data sources such as multimedia content (images, videos) and metadata (timestamps, user interactions) could enrich the analysis. Combining textual data with

other forms of social media content may uncover new patterns and enhance the detection accuracy of depressive symptoms. This multimodal approach could provide a more holistic view of an individual's online behavior and mental health status.

Thirdly, exploring the interpretability of machine learning models, especially deep learning models like LSTM and Bi-LSTM, is crucial. Developing techniques to make these models more transparent and understandable would help in gaining the trust of clinicians and end-users. Explainable AI methods could be applied to highlight which features or aspects of social media activity are most indicative of depression, thereby aiding in clinical decision-making.

Moreover, longitudinal studies tracking changes in social media behavior over time could offer insights into the progression of depressive symptoms. Such studies would help in identifying early warning signs and understanding how online behaviors correlate with the onset and progression of depression. This temporal analysis could enhance the predictive capabilities of machine learning models.

Another important avenue for future research is the ethical consideration of using social media data for mental health monitoring. Ensuring user privacy, obtaining informed consent, and addressing the potential for algorithmic biases are critical components that need thorough exploration. Developing ethical guidelines and frameworks will be essential for the responsible deployment of these technologies.

Lastly, collaboration with mental health professionals to validate and refine the machine learning models could bridge the gap between technical research and practical application. Engaging with clinicians could ensure that the models are clinically relevant and that the insights generated are actionable and beneficial for mental health interventions.

While my project demonstrates the potential of using machine learning for depression detection from social media activity, further studies are necessary to expand the dataset, integrate additional data sources, improve model interpretability, conduct longitudinal analyses, address ethical considerations, and collaborate with healthcare professionals. These steps will enhance the accuracy, applicability, and ethical deployment of machine learning models in mental health monitoring, ultimately contributing to more effective and timely interventions for individuals at risk of depression.

REFERENCE

[1] D. B. Victor, J. Kawsher, M. S. Labib, and S. Latif, “Machine Learning Techniques for Depression Analysis on Social Media- Case Study on Bengali Community,” 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, Nov. 05, 2020. doi: 10.1109/iceca49313.2020.9297436.

[2] Md. Rafidul Hasan Khan, U. S. Afroz, A. K. M. Masum, S. Abujar, and S. A. Hossain, “A Deep Learning Approach to Detect Depression from Bengali Text,” Advances in Intelligent Systems and Computing. Springer Nature Singapore, pp. 777–785, 2021. doi: 10.1007/978-981-33-4367-2_74.

[3] T. F. Mumu, I. J. Munni, and A. K. Das, “Depressed People Detection from Bangla Social Media Status using LSTM and CNN Approach,” Journal of Engineering Advancements, vol. 2, no. 01. SciEnPG, pp. 41–47, Mar. 06, 2021. doi: 10.38032/jea.2021.01.006.

[4] M. K. Kabir, M. Islam, A. N. B. Kabir, A. Haque, and M. K. Rhaman, “Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques,” JMIR Formative Research, vol. 6, no. 9. JMIR Publications Inc., p. e36118, Sep. 28, 2022. doi: 10.2196/36118.

[5] Md. R. H. Khan, U. S. Afroz, A. K. M. Masum, S. Abujar, and S. A. Hossain, “Sentiment Analysis from Bengali Depression Dataset using Machine Learning,” 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, Jul. 2020. doi: 10.1109/icccnt49239.2020.9225511.

[6] T. Ghosh, Md. H. A. Banna, Md. J. A. Nahian, M. N. Uddin, M. S. Kaiser, and M. Mahmud, “An attention-based hybrid architecture with explainability for depressive social media text detection in Bangla,” Expert Systems with Applications, vol. 213. Elsevier BV, p. 119007, Mar. 2023. doi: 10.1016/j.eswa.2022.119007.

[7] A. H. Uddin, D. Bapery, and A. S. Mohammad Arif, “Depression Analysis of Bangla Social Media Data using Gated Recurrent Neural Network,” 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). IEEE, May 2019. doi: 10.1109/icasert.2019.8934455.

[8] M. B. Mohammed, A. S. Md. Abir, L. Salsabil, M. Shahriar, and A. Fahmin, “Depression Analysis from Social Media Data in Bangla Language: An Ensemble Approach,” 2021 Emerging Technology in Computing, Communication and Electronics (ETCCE). IEEE, Dec. 21, 2021. doi: 10.1109/etcce54784.2021.9689887.

[9] Md. N. Hoque and U. Salma, “Detecting Level of Depression from Social Media Posts for the Low-resource Bengali Language,” Journal of Engineering Advancements, vol. 4, no. 02. SciEnPG, pp. 49–56, Jun. 28, 2023. doi: 10.38032/jea.2023.02.003.

[10] M. Billah and E. Hassan, “Depression Detection from Bangla Facebook Status using Machine Learning Approach,” International Journal of Computer Applications, vol. 178, no. 43. Foundation of Computer Science, pp. 9–14, Aug. 19, 2019. doi: 10.5120/ijca2019919314.

[11] S. Islam, Md. J. Islam, Md. Mahadi Hasan, S. M. S. M. Ayon, and S. S. Hasan, “Sentiment Analysis From Bengali Social Media Posts Using Hybridized Feature Extraction Approach,” 2022 IEEE Symposium on Industrial Electronics & Applications (ISIEA). IEEE, Jul. 16, 2022. doi: 10.1109/isiea54517.2022.9873668.

[12] A. K. Das, A. Al Asif, A. Paul, and Md. N. Hossain, “Bangla hate speech detection on social media using attention-based recurrent neural network,” Journal of Intelligent Systems, vol. 30, no. 1. Walter de Gruyter GmbH, pp. 578–591, Jan. 01, 2021. doi: 10.1515/jisys-2020-0060.

[13] A. Amanat et al., “Deep Learning for Depression Detection from Textual Data,” Electronics, vol. 11, no. 5. MDPI AG, p. 676, Feb. 23, 2022. doi: 10.3390/electronics11050676.

[14] S. Alghowinem, T. Gedeon, R. Goecke, J. F. Cohn, and G. Parker, “Interpretation of Depression Detection Models via Feature Selection Methods,” IEEE Transactions on Affective Computing, vol. 14, no. 1. Institute of Electrical and Electronics Engineers (IEEE), pp. 133–152, Jan. 01, 2023. doi: 10.1109/taffc.2020.3035535.

[15] H. Cai et al., “A Pervasive Approach to EEG-Based Depression Detection,” Complexity, vol. 2018. Hindawi Limited, pp. 1–13, 2018. doi: 10.1155/2018/5238028.

[16] A. Husseini Orabi, P. Buddhitha, M. Husseini Orabi, and D. Inkpen, “Deep Learning for Depression Detection of Twitter Users,” Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. Association for Computational Linguistics, 2018. doi: 10.18653/v1/w18-0609.

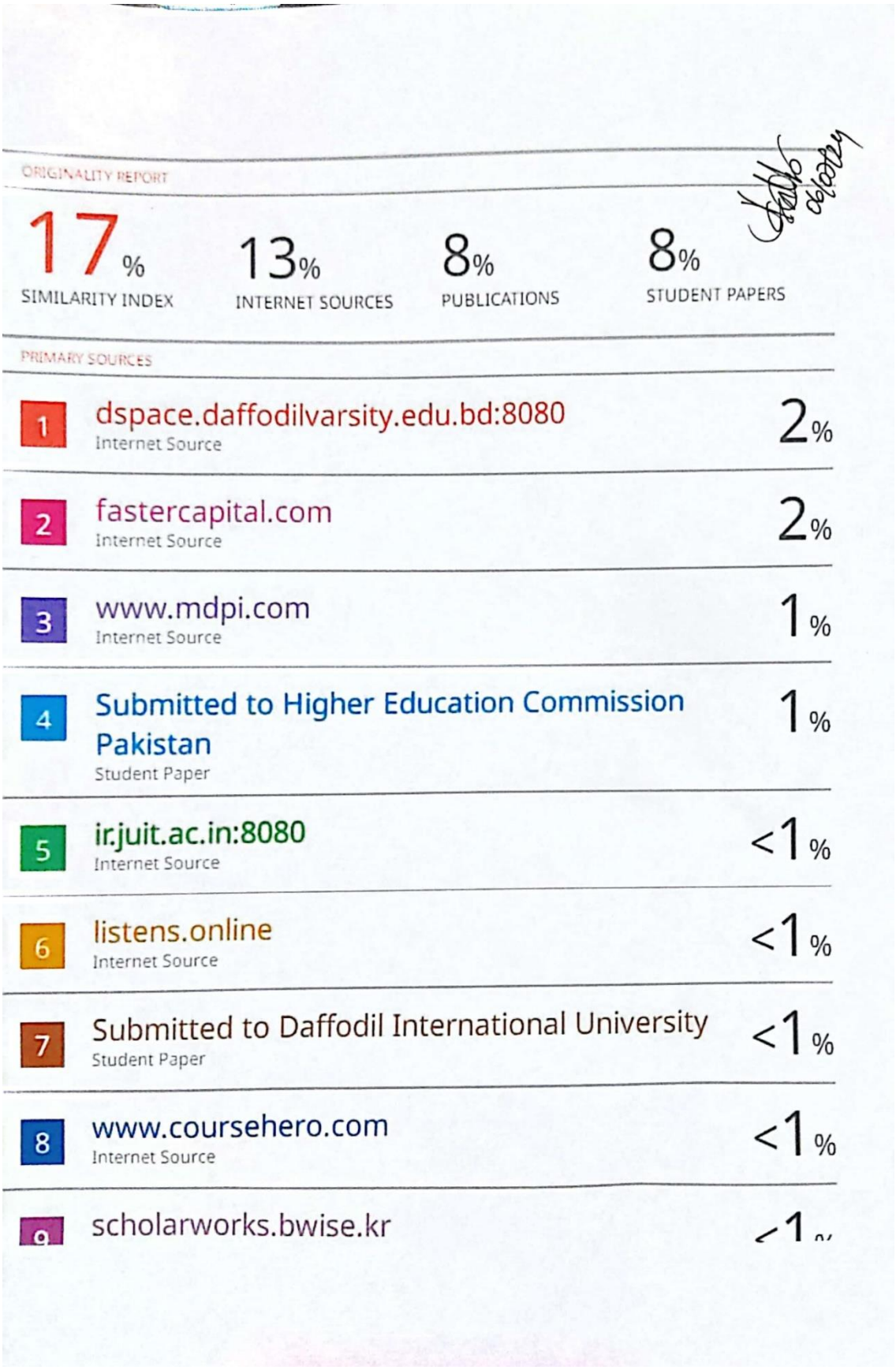
[17] Md. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, “Depression detection from social network data using machine learning techniques,” Health Information Science and Systems, vol. 6, no. 1. Springer Science and Business Media LLC, Aug. 27, 2018. doi: 10.1007/s13755-018-0046-0.

[18] M. Deshpande and V. Rao, “Depression detection using emotion artificial intelligence,” 2017 International Conference on Intelligent Sustainable Systems (ICISS). IEEE, Dec. 2017. doi: 10.1109/iss1.2017.8389299.

[19] C. Lin et al., “SenseMood: Depression Detection on Social Media,” Proceedings of the 2020 International Conference on Multimedia Retrieval. ACM, Jun. 08, 2020. doi: 10.1145/3372278.3391932.

[20] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, and M. Breakspear, “Eye movement analysis for depression detection,” 2013 IEEE International Conference on Image Processing. IEEE, Sep. 2013. doi: 10.1109/icip.2013.6738869.

[21]F. Sadeque, D. Xu, and S. Bethard, “Measuring the Latency of Depression Detection in Social Media,” Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. ACM, Feb. 02, 2018. doi: 10.1145/3159652.3159725.
[22] R. Chiong, G. S. Budhi, S. Dhakal, and F. Chiong, “A textual-based featuring approach for depression detection using machine learning classifiers and social media texts,” Computers in Biology and Medicine, vol. 135. Elsevier BV, p. 104499, Aug. 2021. doi: 10.1016/j.compbiomed.2021.104499.



ORIGINALITY REPORT

17%	13%	8%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
2	fastercapital.com Internet Source	2%
3	www.mdpi.com Internet Source	1%
4	Submitted to Higher Education Commission Pakistan Student Paper	1%
5	ir.juit.ac.in:8080 Internet Source	<1%
6	listens.online Internet Source	<1%
7	Submitted to Daffodil International University Student Paper	<1%
8	www.coursehero.com Internet Source	<1%
9	scholarworks.bwise.kr	<1%