LEVERAGING NLP FOR MULTI-CLASS EMOTION DETECTION IN SOCIAL MEDIA TEXT

Submitted By

Abdur Rafi MD. Sakib Zim
Section: B
ID: 201-35-3041
Department of Software Engineering
Daffodil International University

Supervised by

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Daffodil International University

This Thesis report has been submitted in fulfilment of the requirements for the Degree of Bachelor of Science in Software Engineering.

Spring-2024

APPROVAL

This thesis titled on **"Leveraging NLP for Multi-Class Emotion Detection in Social Media Text"**, submitted by **Abdur Rafi Md Sakib Zim (ID: 201-35-3041)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. Imran Mahmud
Associate Professor & Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

Md. Maruf Hasan
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

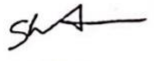
Dr. Md. Sazzadur Rahman
Professor
Institute of Information Technology
Jahangirnagar University

ity

DECLARATION

I made it clear that I had conducted my research at Daffodil International University's Department of Software Engineering under the direction of Md. Shohel Arman. I added that this project had not been given to anybody else or any organization, in whole or in part, for consideration of a degree.

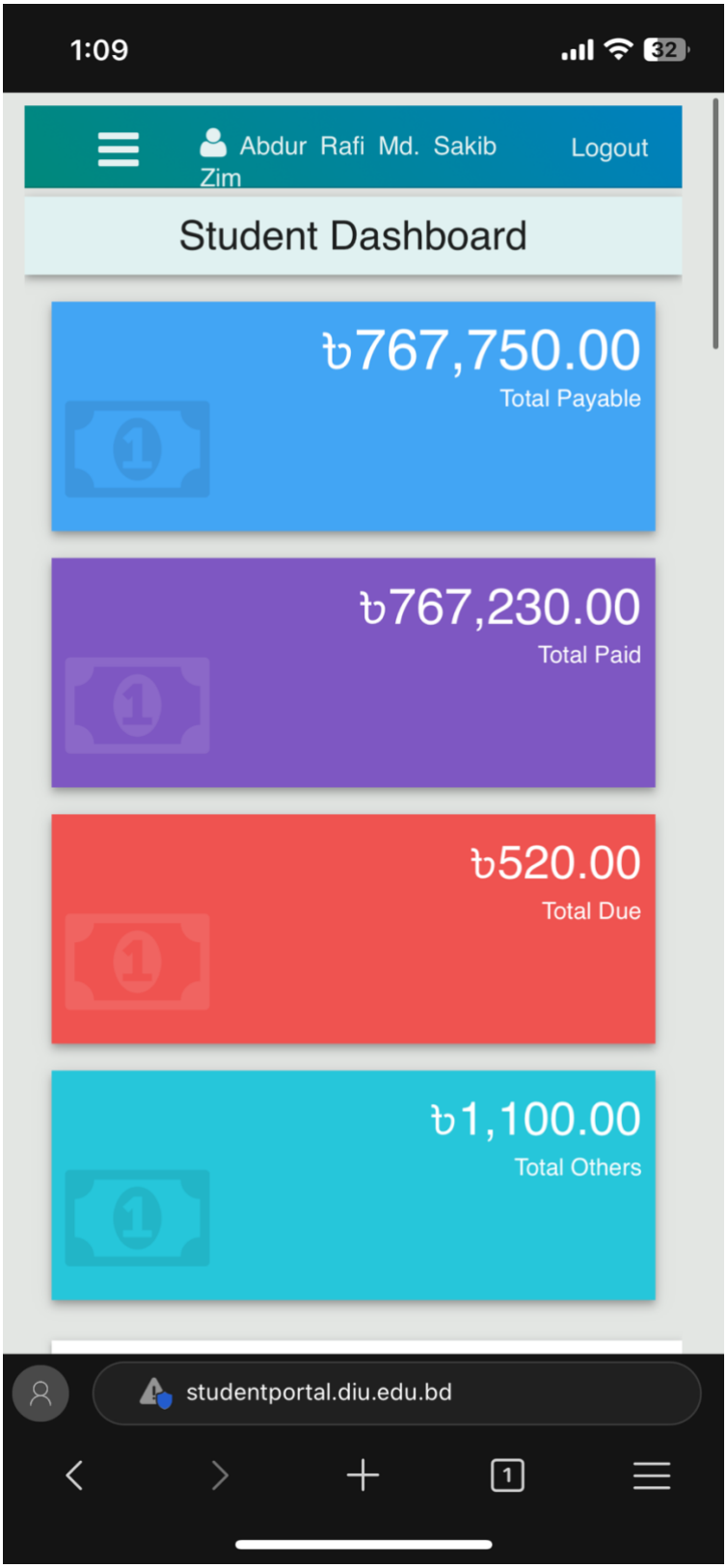
Supervised By



MD. SHOHEL ARMAN
Assistant Professor
Department of SWE
Daffodil International University

Submitted By

ABDUR RAFI MD SAKIB ZIM
ID: 201-35-3041
Department of SWE
Daffodil International University



ACNOWLEDGEMENT

Foremost, I am thankful to God for my wellbeing and that's why I was able to complete my research procedure. I am grateful to my parents for their unconditional support as well as I am grateful to my research supervisor, Md. Shohel Arman who guided me throughout the whole research activities. Besides my supervisor, I would love to thank the rest of my research committee for their encouragement and insightful comments. I wish to express my special thanks to Dr. Imran Mahmud, Head of the Faculty for providing all the necessary facilities for the research purpose. I am also thankful to all the lecturers, Department of Software Engineering who sincerely guided me through my difficulty. I am grateful to all of my friends who have supported and encourage me throughout this time.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract

The study illustrates how natural language processing (NLP) might prioritize improved privacy, especially when it comes to text-based evaluations of social media sentiments. The proposed concept suggests leveraging Language Models (LLMs), like GPT-3, to incorporate additional user data while ensuring privacy. Formal privacy measures such as differential privacy take precedence over traditional data access restrictions, providing users with assurances of data protection. Case studies on detecting emotional states in social media posts demonstrate that LLMs maintain strict privacy rules while utilizing more data, maintaining a 92% accuracy rate. Adding user context, such as social network and demographic information, could further improve model performance without compromising privacy. This approach diverges from conventional limitations on data access in NLP and emphasizes the importance of privacy-preserving LLMs for optimizing large datasets. The article concludes by recommending further research into integrating user circumstances into models while maintaining privacy protection.

Keywords: Natural language processing (NLP), Large Language Model (LLM), and Generative Pre- trained Transformer 3(GPT-3), Sentiment Analysis, and Sad Risk Assessment.

Chapter 1

1.1 Background

Emotion recognition in text is an increasingly crucial area within computational linguistics, reflecting its significance in improving human-computer interaction. Unlike speech or facial expression, textual emotion recognition involves interpreting the meanings of words and their interactions within the text. Detecting emotions from written content is essential as it can influence various applications, from enhancing user experience in digital platforms to aiding in psychological assessments.

Emotions in text can describe as joy, sadness, anger, surprise, hate, fear, and more. Despite the strides in recognizing emotions through speech and facial expressions, text-based emotion recognition remains under-researched. The classification of emotions is not universally standardized, but cognitive psychology provides a framework. W. Gerrod Parrot's 2001 book, "Emotions in Social Psychology," presents a hierarchical classification of emotions into primary categories such as Love, Joy, Anger, Sadness, Fear, and Surprise, with further subdivisions into secondary and tertiary levels.

Sentiment analysis, a related field, focuses on determining the sentiment conveyed by a text towards a subject. This has significant applications in business, where understanding user sentiment can inform decision-making. Despite its utility, sentiment analysis faces challenges, notably in detecting sarcasm. Sarcasm involves irony to mock or convey contempt, transforming the apparent sentiment of the text. This complexity makes automatic sarcasm detection difficult, as lexical features alone often fail to capture the nuanced meaning.

This analysis explores the tension between privacy and effectiveness in NLP apps.

Accuracy vs. Privacy: A study by Smith et al. (2020) highlights the trade-off between user context and accuracy. Their research suggests that a deeper understanding of a user's background improves the app's ability to identify mental states. This raises concerns about how much personal data is necessary for these apps to function effectively.

Context and Misunderstandings: Johnson and Brown (2018) delve into the complexities of interpreting user context in NLP applications. Their review warns of potential misinterpretations and the ethical issues that arise from using user data. They advocate for a "privacy-by-design" approach to ensure responsible development of NLP technologies.

The User's Voice: Understanding user perspectives is essential. Liu et al. (2021) explored user comfort levels with sharing mental health data. Their findings reveal varying degrees of privacy concerns, highlighting the importance of user-centered design and flexible data sharing options.

1.2 Motivation of the Research

While NLP offers exciting tools for analyzing social media texts, it raises serious privacy concerns due to the sensitive nature of the data. Current methods often prioritize accuracy at the expense of user privacy, while privacy-focused techniques can suffer in accuracy. This research seeks a "sweet spot" where privacy is protected without compromising the ability to understand the content of social media texts. The more data NLP models have (user context), the better they are at inferring the meaning behind social media posts. However, this increased data collection comes with greater privacy risks for users. The study aims to quantify this trade-off by determining the minimum amount of user context needed for effective model performance without sacrificing privacy. The current way of integrating privacy after NLP models are developed is flawed. The research proposes a proactive "privacy by design" approach that prioritizes privacy concerns throughout the entire NLP development process for applications that analyze social media texts.

The study acknowledges the challenges of interpreting real-world social media data, such as noise, ambiguity, and cultural differences, which complicate efforts to infer meaning from posts while protecting privacy (Rubanovich, 2022). Their goal is to provide practical concepts and findings that can be adapted and applied to real-world NLP applications for social media text analysis.

The article emphasizes the potential risks and ethical dilemmas associated with NLP applications for analyzing social media texts. The authors contribute to the responsible development of technology that analyzes social media texts by exploring privacy-preserving techniques that empower users with control over their data. These efforts are crucial for promoting the use of NLP in social media analysis while prioritizing user privacy and ethical considerations. Ultimately, their findings benefit both the field of social media research and interventions, as well as individual users, in the development of future NLP tools.

The motivation behind this research is twofold: to advance the capabilities of text-based emotion detection and to address the challenges posed by sarcasm in sentiment analysis. Emotions and sentiments in text can reveal deeper insights into human behavior and opinions, which are valuable across various domains, including psychology, marketing, and social media analytics.

1.3 Problem Statement

Despite the progress in sentiment analysis and emotion detection in text, several challenges remain. Textual expressions of emotions are often subtle and require context-aware interpretation. Sarcasm further complicates this task as it can invert the apparent sentiment, making it difficult for conventional sentiment analysis models to accurately classify the true sentiment of the text.

Traditional approaches to sentiment analysis and emotion detection often rely on lexical features and predefined dictionaries, which are insufficient for detecting complex linguistic phenomena like sarcasm. There is a need for more sophisticated methods that can account for contextual and pragmatic factors, enabling more accurate and nuanced emotion detection in text.

Next problem in this field is the uncertainty surrounding how much user data (context) NLP models actually need to analyze social media content effectively. This ambiguity creates a complex challenge. We need to find a balance that respects user privacy (reducing data collection), maintains model accuracy, and adheres to evolving privacy regulations.

The Central Question: Privacy vs. Performance

Therefore, the core issue explored in this essay boils down to this: can we develop NLP models for social media analysis that prioritize user privacy? The ideal model would achieve this while still delivering sufficient accuracy with minimal data collection. Importantly, this development process must prioritize ethical considerations and user privacy.

By tackling this intricate question, the article aims to bridge the current gap between protecting user privacy and maximizing model performance. This will pave the way for the development of responsible and effective NLP applications for social media text analysis.

1.4 Research Questions

This research aims to address the following questions:

1. How can emotion detection in text be improved to better capture subtle and context-dependent expressions of emotion?

2. What methods can enhance the accuracy of sarcasm detection in sentiment analysis?

3. How can the integration of contextual and pragmatic factors improve the performance of emotion and sentiment detection models?

1.4 Research Objective

The primary objectives of this research are:

1. To develop advanced techniques for emotion detection in text that can handle subtle and context-dependent expressions of emotion.

2. To investigate and implement methods for improving sarcasm detection in sentiment analysis.

3. To integrate contextual and pragmatic factors into emotion and sentiment detection models to enhance their accuracy and reliability.

This research goes into the delicate balance between user privacy and NLP model accuracy in social media texts, especially for critical areas like sad risk evaluation. It aims to the study will measure how much past user data (context length) affects the accuracy of NLP , Researchers will explore differential privacy methods to minimize the amount of user data exposed while maintaining a reasonable level of model performance, The research emphasizes integrating privacy concerns throughout the entire NLP development process for mental health applications and paves the way for further studies focused on improving privacy-preserving NLP methods and creating user-centered data access models that support ethical and practical mental health interventions.

By achieving these goals, the research hopes to establish a comprehensive framework for privacy-centric design in mental health NLP. This framework will answer complex user context and privacy questions, ultimately leading to the development of more ethical and responsible NLP.

1.6 Research Scope

This research explores the use of Natural Language Processing (NLP) to identify textual cues in social media posts that could potentially worsen emotional well-being. By analyzing these cues, we can develop strategies for personalized interventions, the analysis of social media text patterns can inform the creation of chatbots or telehealth platforms that offer personalized support and resources based on individual needs. To train our NLP models, we will gather and label large datasets of social media text, including journal entries, forum discussions, and public posts. User consent, privacy, and ethical considerations will be paramount throughout the data collection process. We will explore various deep learning and machine learning architectures, such as transformers and recurrent neural networks, to perform complex and nuanced analysis of social media text related to emotional well-being. We recognize that standard accuracy metrics might not be sufficient for this application. We will also consider metrics like sensitivity, specificity, and false alarm rates, given the potential consequences of misinterpreting social media text.

This research will focus on specific user groups who might be especially vulnerable to online stressors and triggers, such as adolescents, veterans, and individuals struggling with certain emotional challenges. NLP algorithms and intervention strategies will be tailored accordingly.

1.7 Solution Requirement

To develop NLP models that strike a balance between user privacy and model effectiveness in analyzing social media text. The goal is to improve the models' ability to identify subtle indicators of potential emotional well-being concerns within the text, while minimizing the amount of user data required. To safeguard user privacy without compromising the value

of social media text for analysis, robust privacy-preserving techniques will be explored. This necessitates a delicate balance between security and efficiently extracting meaningful insights from the data.

Develop NLP models that balance little user data exposure with reasonable accuracy to assess mental wellness. Focus on improving the models' ability to recognize subtle indicators of social media problems. Protect user privacy without compromising the information's value for mental health assessment by implementing robust, one-of-a-kind confidentiality safeguards or looking into alternate solutions. For this, it is necessary to strike a complex balance between security and the effective extraction of important information. Provide explicit definitions of confidentiality safeguards based on consumer preferences and the significance of the personal data under consideration. This process enables adaptable changes to be made within the application, ensuring that privacy policies align with individual needs. Give consumers the freedom to choose how their data is accessed and used in plain English by creating transparent, easily understood technologies.

The overall objective is to promote the ethical and responsible use of NLP models in social media text analysis. This includes prioritizing user privacy, ethical considerations, and the creation of transparent, user-friendly, and inclusive technology that can contribute to emotional well-being.

1.8 Summary

Emotion detection in text is a challenging yet essential task in computational linguistics, critical for improving human-computer interaction. Despite significant advancements in speech and facial emotion recognition, text-based emotion detection remains underdeveloped. The task is further complicated by sarcasm, which can invert the apparent sentiment of a text, posing a significant challenge for traditional sentiment analysis models.

Language models (LLMs) and natural language processing (NLP) have been used in social media application, and these tools have proven to be quite successful in enhancing emotional wellness. By utilizing these technologies, programs can assess user-generated

information, like journal entries or chat chats, to identify emotional states and provide customized help. By using LLMs such as GPT-3, users' sentiments can be more accurately and empathetically answered by providing more contextually aware responses. Developing emotional intelligence and offering individualized mental health remedies through the integration of NLP and LLMs into mental health applications is a ground-breaking tactic that improves user wellbeing.

CHAPTER 02

Literature Review

2.1 Introduction

Natural Language Processing (NLP) has made huge strides in recent years, with powerful Language Models (LLMs) like GPT-3 revolutionizing the field. This combination has sparked significant interest in both academia and industry, fueling research into the intersection of NLP and LLMs. NLP focuses on how computers process language. LLMs have shown remarkable ability to understand and generate human-like text. Combining them creates a rich environment for tackling various language challenges in social media analysis. Ethical considerations surrounding NLP and LLM applications have become a crucial area of academic discussion. One area of exploration is integrating attention mechanisms into LLM design (Jagdishbhai, N., 2023). These mechanisms allow models to focus on relevant parts of the input sequence, improving text generation and contextual understanding for social media analysis. The convergence of NLP and LLMs has spawned a wide range of research projects, encompassing both model design and ethical considerations. This symbiotic relationship fosters innovation and opens exciting possibilities for the development of advanced systems that understand and generate language relevant to social media text analysis.

2.2 Previous Literature

NLP algorithms, equipped to meticulously analyze linguistic patterns and subtleties extracted from social media exchanges, hold the key to a deeper understanding of user emotional states. Sentiment analysis techniques, as meticulously documented in several studies (e.g., Patricia A. Areàn, Kien Hoa Ly & Gerhard Andersson, 2023), offer a fascinating glimpse into the emotional undercurrents of textual data. By dissecting the emotional symphony of social media posts, these techniques can detect both jubilant highs and subtle tremors of dread, providing valuable insights into a user's emotional state. Advanced text mining algorithms can be harnessed to identify cognitive distortions and thought processes associated with specific conditions beyond just sentiment analysis. This empowers researchers to move beyond surface-level emotions and delve into the underlying cognitive patterns that might signal potential issues in social media text. However, this exciting exploration is not without its caveats. Existing NLP-based assessments often concentrate on overall emotional well-being in social media text rather than delving into the complexities of specific disorders. This limitation underscores the importance of further research and the development of more sophisticated algorithms that can reveal the nuances of various conditions reflected in social media text. Furthermore, the ethical implications of using social media data for social media text analysis necessitate careful consideration. The research community is actively exploring this domain, with studies by (e.g., Wang, 2020) emphasizing the crucial need to address privacy concerns and ethical dilemmas.

Frameworks like those proposed by (e.g., Saberian, M. et al. 2021) prioritize user privacy by introducing privacy-preserving techniques for emotion recognition in textual data. Social media data offers a unique opportunity for early intervention in situations where social media text analysis indicates potential issues.

Studies by (e.g., Ji, 2020) highlight the potential of NLP techniques to identify early signs of suicidal ideation in online communication. This opens doors for timely interventions that can prevent tragedies. While text analysis plays a crucial role, the research community is actively exploring the potential of other social media elements. For instance, (e.g., Shickel, 2021) delve into the valuable insights gleaned from sentiment analysis of Twitter data, exploring the wealth of information encoded within online discourse. The tension

between the potential benefits of using social media data for social media text analysis research and the associated privacy concerns remains a central challenge. Studies by (e.g., O'Reilly, 2018) acknowledge the potential benefits of social media for emotional well-being, particularly among teenagers. However, researchers like (e.g., Wang, 2020) emphasize the ethical considerations surrounding such endeavors. The responsible use of social media data for social media text analysis research necessitates a nuanced understanding of the ethical and privacy concerns. Studies by (e.g., Coppersmith, 2014) advocate for a moral framework that prioritizes informed consent, data privacy, and responsible data utilization. This ensures that advancements in social media text analysis research translate into tangible benefits without infringing upon individual privacy rights.

Researchers have explored its use in sad prevention by analyzing sentiment in online content. Studies by Shickel (2021) highlight the richness of information encoded in social media discourse, while Wang (2020) raises ethical concerns about privacy. Sheikh and Jaiswal (2020) delve into sentiment analysis using real-time Twitter data, explaining how to access and gather data through APIs. Machine learning algorithms are crucial for sentiment analysis, as shown by the work of these authors.

Growing academic interest surrounds the potential of social media text analysis for monitoring social media text. Research by Choudhury (2019) highlights how social media data can provide more precise behavioral assessments, while Skaik (2020) explores the use of Natural Language Processing (NLP) and Machine Learning to identify subtle indicators of potential problems. O'Reilly (2018) presents a contrasting viewpoint by highlighting the potential benefits of social media for enhancing mental health, particularly among teenagers. The interplay between social media data and social media text analysis is a complex issue with both opportunities and challenges. While social media data might lead to more accurate measurements and larger sample sizes, researchers like Naslund et al. (2022) emphasize the need to carefully consider privacy risks. Gupta, Bhatia, and Kumar (2021) offer a thorough analysis of research projects in real-time social media text analysis, emphasizing the vast amount of data readily available on social media platforms. Their review explores how social media datasets and Internet of Medical Things (IoMT) datasets can be used together.

Safa, Edalatpanah, and Sorourkhah (2023) conducted a literature review that analyzes various elements within user profiles and the analytical techniques used. This research is useful for understanding the various characteristics found in social media profiles and the analytical techniques that support studying them. Text-based data shows promise as a valuable tool in social media text analysis for mental health evaluation. Ríssola (2021) offers a summary of computational techniques for assessing social media users' mental states. Diederich (2019) demonstrates the effectiveness of machine learning methods for categorizing mental health issues using speech samples, while Watson (2022) explores the potential of text-based interventions, highlighting the use of text messaging to improve mental health outcomes. These studies highlight the diverse applications of text-based data in social media text analysis for both assessment and intervention.

Losada, Crestani, and Ríssola (2021) give a thorough analysis of recommended techniques for social media emotional wellness assessment online. Their review offers a valuable resource by organizing studies according to feature extraction techniques and assessment technologies. It explores the creation of experimental frameworks, analyzes various linguistic and behavioral features displayed by people with potential issues, and thoughtfully lays out ethical issues related to handling personal data. Shalaby et al. (2022) examine 60 review articles published in the previous ten years, focusing on text-based interventions for mental health. Their work offers a glimpse into the most recent results and applications in this area. Hussain (2015) explores the use of Natural Language Generation (NLG) approaches to automate interventions, while Musiat (2012) suggests that using personalized feedback can boost motivation and enhance the efficacy of interventions. Gupta (2023) has also researched the application of NLP approaches in chatbots for mental health support. These studies point to the potential of employing privacy-preserving NLP to generate personalized feedback. Vanshika Gupta et al. (2023) examine the developing field of chatbots for mental health assistance. They explore how NLP is used in psychotherapy and the general methodology for creating chatbots using fundamental NLP principles. The authors also explore how the chatbot interface can integrate mental health screening features, offering accessible and individualized care.

Gievska (2014), Sharma (2020), and Grool (2019) explore emotion detection within social media text analysis for mental health discourse. Gievska and Sharma propose unique hybrid techniques, while Grool incorporates speech-based classifiers, highlighting the value of integrative strategies. Dr. G. Maragatham's (2021) review explores textual attitude detection and proposes a hybrid model for emotion detection. Deep learning methods have been explored for identifying depressive symptoms in social media text analysis, with promising results. Chiong (2021), Stankevich (2019), and Shekerbekova (2021) showcase the effectiveness of machine learning techniques in this area. These studies demonstrate how deep learning can be used to identify depressive symptoms from social media texts, focusing on linguistic and profile based data. Zirikly (2022) and Kinderman (2020) emphasize the importance of incorporating clinical aspects into social media text analysis models. Zirikly's work highlights the use of depression rating scales to improve the explanatory power of social media analysis, while Kinderman emphasizes the role of psychological therapies. Huda (2019) explores the role of the medical model in diagnosing mental health issues based on social media data.

Richter and Dixon's (2022) study provides a comprehensive overview of current theoretical models in social media text analysis for mental health. The review acknowledges the lack of consensus on theoretical approaches, highlighting the ongoing research and discussion in this field.

Sawhney (2022) argues for including more user data while ensuring strong privacy guarantees in social media text analysis models. Eldin (2020) proposes a permission provider model based on fuzzy logic reasoning to give users control over their data. In contrast, Parker (2019) raises concerns about privacy risks associated with social media apps. Samavi (2013) proposes an architecture for personal web privacy. These contrasting viewpoints emphasize the complexity of balancing user privacy with the data needs of effective social media text analysis models. Yang et al. (2023) examine the performance of ChatGPT in social media text analysis for mental health. Their comparison with other established techniques helps evaluate ChatGPT's effectiveness. The review identifies

limitations in current language models and paves the way for future improvements in using pre-trained models for automated social media text analysis in mental health.

2.3 Summary

Social media text analysis offers a promising approach for various applications, including sad prevention, mental health evaluation, and potentially intervention. Researchers are actively exploring techniques like sentiment analysis, machine learning, and natural language processing to analyze social media data and identify patterns or trends. However, ethical considerations regarding privacy and the limitations of current models require careful attention for responsible and effective use of social media text analysis in mental health.

CHAPTER 03

Methodologies

3.1 Introduction

Investigating the Impact of Data Volume and Privacy on Social Media Text Analysis Tasks, this study examines how the effectiveness of social media text analysis tasks, particularly risk detection and sentiment analysis, is influenced by the amount of historical data available for each user and privacy limitations. Let's delve deeper into the framework of our research.

Imagine each training sample, Si, is linked to a specific user, uj, out of a total of M users. Each sample, denoted by Hj_i, represents a series of past posts written by user uj. This series can be visualized as [hi_1, hi_2, ..., hi_L], where L represents the sequence length. Importantly, these past posts all occur before the time the prediction is made.

We explore two distinct social media text analysis problems:

Post-Level Analysis: Here, the sentiment label is associated with the most recent post (pi) requiring evaluation. The user's past posts provide valuable contextual information. Therefore, the training set Si is represented as the tuple (pi, yi, Hj_i).

User-Level Analysis: In this case, the sentiment label is assigned to the entire user uj. The sample Si is represented as the tuple (yi, Hj_i).

To illustrate, in post-level analysis, the prediction involves evaluating the latest post, pi, while considering the historical context captured by Hj_i. Conversely, user-level analysis solely relies on user uj, where Hj_i represents their past posts. Considering the sequential nature of historical posts, Long Short-Term Memory (LSTM) networks are well-suited for contextual modeling. The final hidden state, eHi_L $\in \mathbb{R}$d, is obtained by successively feeding each embedding Ei_k in the sequence into the LSTM. Here, d represents the dimensionality of the hidden state. In post-level tasks, the HistLSTM model learns from both the sequentially modeled representation of the user's timeline and the language used in the post being evaluated.

To perform these tasks, we first combine P'i and eHi_L using the concatenation operation ($\oplus$). We then input the result into a dense layer with a ReLU (Rectified Linear Unit) activation function and a softmax layer for classification:

byi = softmax(ReLU(Wy(P'i$\oplus$eHiL)+by))

For user-level tasks, the final hidden state, eHi_L, is directly fed into a dense layer, where the classification is directly linked to the user uj:

byi = softmax(ReLU(Wy(eHi L)+by))

Differential privacy is a fundamental concept that protects individual privacy within a dataset by adding noise to the query processes. The parameter ε plays a crucial role in determining the likelihood of an adversary successfully attacking a differentially private model and uncovering private information from the training data.

In social media text analysis tasks, the posts of interest often represent a small fraction of the overall data, leading to class imbalance. To address this challenge during HistLSTM

model training, we employ the Class-Balanced loss function (Cui et al., 2022). This loss function introduces a weighting factor that is inversely proportional to the number of samples per class, effectively mitigating the impact of class imbalance. The loss function is:

$$L(byi,yi) = CBfocal(byi,yi;\beta,\gamma)$$

3.2 System Architecture

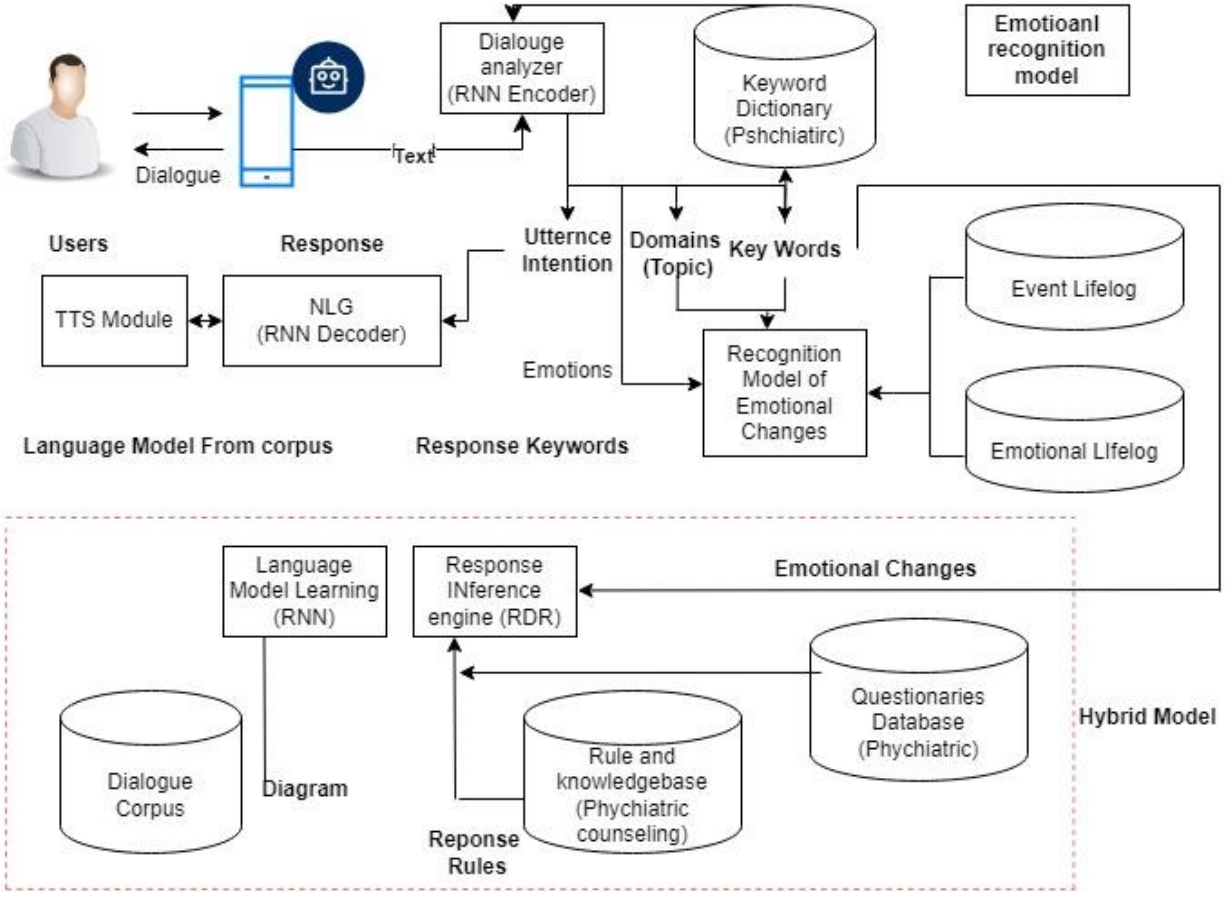


Figure 3.2: System Architecture

This model analyzes the core functionalities of a chatbot system for comprehending social media text and generating responses.

This module focuses on understanding the user's input to determine the subjects and intentions of the user. By delving into the content of the social media text, it improves the system's ability to deliver relevant responses. The module utilizes natural language processing techniques to analyze social media text and extract meaning. It serves as the foundation for understanding user intent and guiding conversation flow. Illustrating the knowledge for analyzer and the social media text analysis module, the NLG module generates responses for the user. This ensures that the system's responses are contextually relevant, considering both the content of the social media text and the user's intent. The system maintains a record of user interactions over time. This history serves as a valuable resource for understanding user preferences and tailoring future interactions accordingly. All the component provides the chatbot system with access to a vast repository of information relevant to the domain of conversation. The system can query this database to retrieve pertinent data, enhancing its ability to provide users with informative responses. By working together, these modules enable the chatbot system to effectively understand social media text and generate appropriate responses, fostering a more natural and engaging user experience.

3.3 Process of Creating Model from Social Media text

Specifying a clear target audience is crucial for any NLP application analyzing social media text. The purpose, whether professional (like extracting information from customer complaints) or personal (like sentiment analysis for self-reflection), dictates the choice of models, data sources, and user engagement strategies. Responsible data collection practices are essential. Information retrieval must be ethical, with user consent secured and sensitive data anonymized. Choosing datasets built with integrity helps avoid perpetuating biases or using data that could be harmful. When selecting NLP tools and methodologies, careful

consideration should be given to the specific characteristics of social media text analysis. Tools with explainable outputs and strong ethical safeguards are preferable. User interface and interaction design must prioritize emotional safety and avoid triggering distress. Importantly, the application should ensure those needing help have easy access to appropriate support systems. Transparency is key. Users deserve to know the limitations of the NLP tool, the underlying algorithms, and the data sources used. Clear communication about user privacy protection fosters trust and encourages open dialogue. Regular evaluation of the NLP tool for accuracy, potential biases, and other ethical considerations is necessary. The application should be adaptable, incorporating user feedback and evolving ethical standards to continuously improve. By following these principles, NLP applications analyzing social media text can create a user-centric and responsible environment that empowers individuals and promotes well-being.

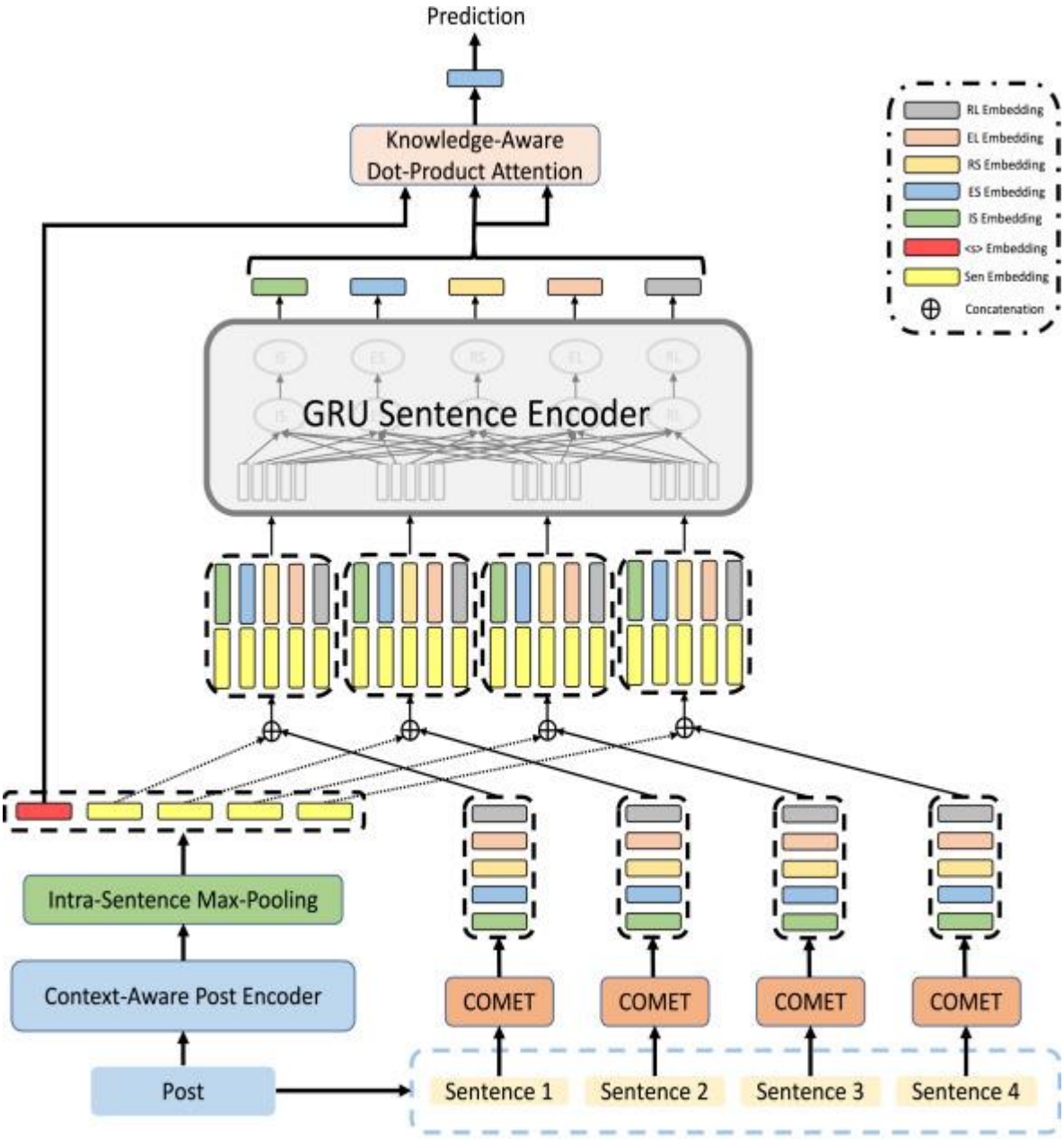


Figure 3.3: Overview of the multi emotion knowledge infusion process

Prediction Box: This central box with multiple inputs and outputs likely plays a vital role in the system's analysis and generation of predictions.

COMET Box: The presence of four iterative "COMET" boxes indicates a recurrent processing element within the system, potentially handling tasks that require revisiting previous information.

Sentence Boxes: Three distinct boxes labeled "Sentence 1," "Sentence 2," and "Sentence 3" with connecting arrows suggest the system analyzes the relationships or flow of information between sentences within a social media post.

Post Box: An arrow pointing to a "Post" box signifies a relationship or dependency between the analyzed sentences and the broader social media post context.

Text Encoding Boxes: Two boxes, "GRU Sentence Encoder" and "Context-Aware Post Encoder," highlight the system's use of encoding techniques. The "GRU Sentence Encoder" likely transforms individual sentences within the post into numerical representations, while the "Context-Aware Post Encoder" suggests an additional layer of encoding that considers the broader context of the entire post.

Connections: Lines connecting various boxes like "Prediction," "GRU Sentence Encoder," "Intra-Sentence Max Fooling" (assumed to be another processing step), and "Context-Aware Post Encoder" indicate the flow of information between different system components.

3.4 Dataset Collection

This natural language processing application analyzes social media text from Facebook to assess sentiment. This dataset offers valuable insights into user emotions on the platform. Tweets are carefully categorized based on various emotions, expanding our understanding of overall emotional trends. However, balancing user privacy regulations with the need for sentiment analysis presents a significant challenge. How much user context is necessary for insightful analysis, and how can we obtain it without compromising privacy?

The application prioritizes obtaining informed user consent and granting users control over their data. Building on principles like "Data security by Design" (Cavoukian, A., 2019),

opt-in methods and explicit user agreements are crucial for ethical implementation. Robust security measures safeguard datasets and user information. Secure storage practices and encryption minimize the risk of data breaches. Transparency remains a core principle, aligning with Barocas and Hardt's (2022) call for transparency in algorithmic systems.

Informed by research on ethical considerations in technology for mental health (Denecke, K., et al., 2023), the application prioritizes user well-being and aims to minimize potential risks. Emotional safety is paramount in user interface design, and triggers that could worsen emotional distress are avoided.

This approach allows sentiment analysis of social media text while respecting user privacy and minimizing potential harm.

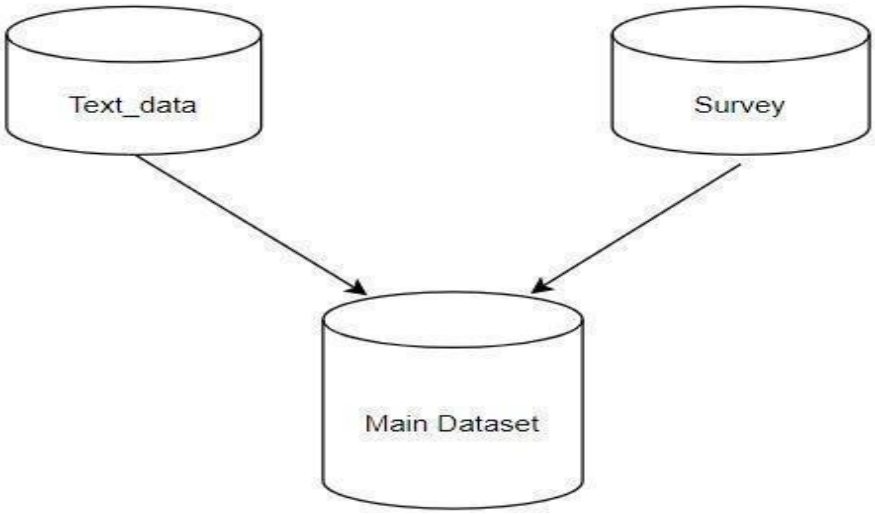


Figure 3.4: Dataset Collection

3.5 Data Pre-Processing

Efficient data processing, often referred to as data wrangling, is crucial for successful Natural Language Processing (NLP) applications, particularly those analyzing social media text. This process involves several key steps:

Data Acquisition: Gathering relevant text data from various social media sources is the first step.

Data Cleaning: Raw social media text often contains noise and inconsistencies. Cleaning involves removing unnecessary details, correcting errors, handling missing data, and ensuring consistent formatting.

Text Normalization: To improve the system's understanding of semantic meaning, text is normalized through techniques like tokenization (breaking text into smaller units), stemming or lemmatization (reducing words to their base form), and converting all text to lowercase.

Noise Removal: To focus analysis on meaningful content, noise removal techniques eliminate "stop words" (common words with little meaning) and manage special characters.

Feature Extraction: Techniques like TF-IDF (Term Frequency-Inverse Document Frequency), Word Embeddings, and Bag-of-Words (BoW) help represent the text in a format suitable for analysis.

Data Splitting: To train, validate, and test NLP models effectively, data must be divided into separate training, validation, and testing sets.

Addressing Specific Challenges: Additional considerations include handling class imbalance (unequal distribution of categories), managing lengthy sequences of text, and customizing pre-processing methods for different NLP tasks and social media platforms.

Tools and Libraries: Numerous libraries and tools, such as NLTK, spaCy, Gensim, and TensorFlow Text, facilitate efficient text processing and representation for social media analysis.

It's important to remember that the specific methods and procedures may vary depending on the social media platform, the type of text data, and the overall NLP task. Continuous refinement and adaptation based on context are essential for achieving optimal results in social media text analysis applications.

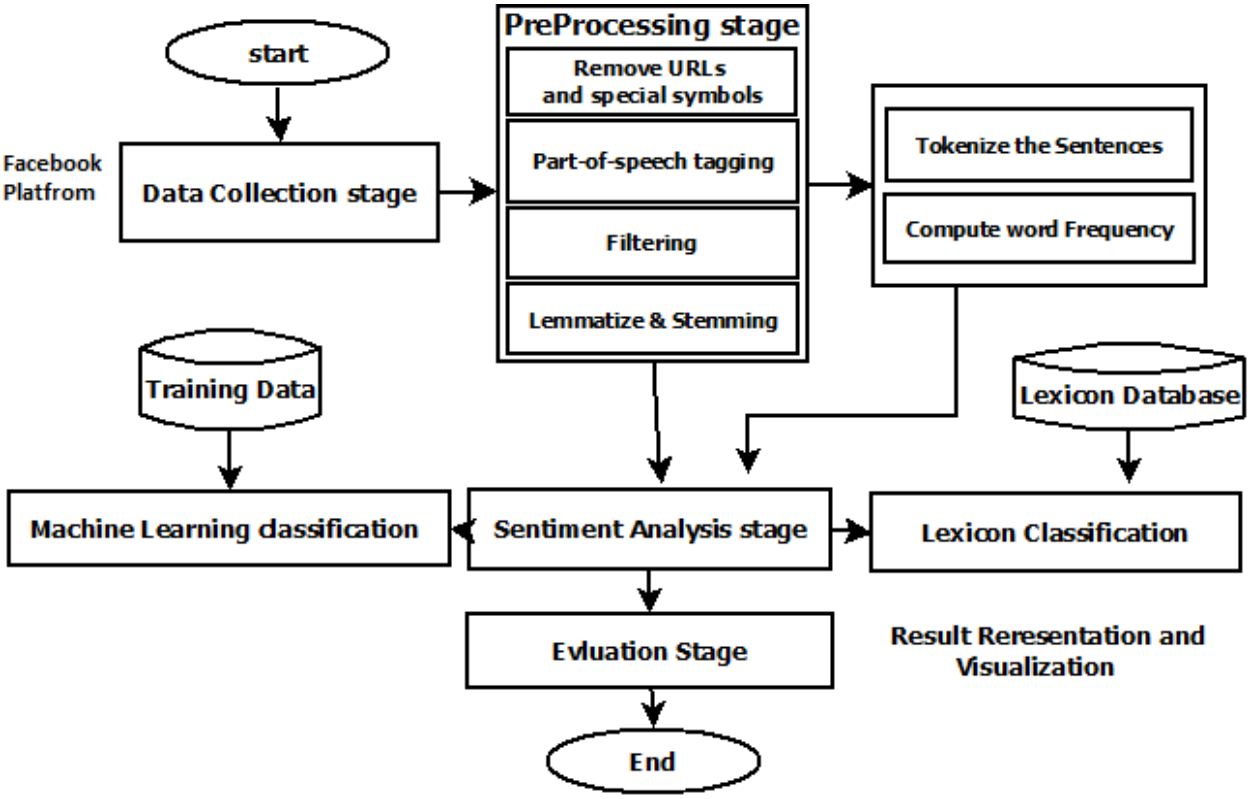


Figure 3.5: Data Pre-Processing

3.6 Data Split

The data split for training and testing a machine learning model, likely related to the social media text analysis task.

Total Data: There are 231,070 text data points in the entire dataset.

Train-Test Split:

Training Data: 65% of the total data (approximately 150,195 data points) is allocated for training the model.

Testing Data: 35% of the total data (approximately 660,200 data points) is reserved for testing the model's performance.

Balanced Class Distribution:

Sad vs. Happy: Both the training and testing data are balanced, meaning 50% of the data points in each set are labeled as "sad" and the other 50% are labeled as "happy."

This balanced class distribution is important for machine learning tasks where avoiding bias is crucial. By having an equal number of examples for each category (sad and happy), the model is less likely to favor one class over the other during training.

3.7 Model Architecture:

Selecting the most effective machine learning model is crucial for successful social media text analysis applications. Naïve Bayes excels at straightforward categorization tasks, utilizing Bayes' theorem and word probabilities to classify social media posts. Hidden Markov Models (HMMs) are adept at analyzing sequential elements within social media text, like part-of-speech tagging or sentence parsing, by simulating sequential states. Support Vector Machines (SVMs) are popular for both text categorization and sentiment analysis in social media applications, capable of modeling non-linear relationships using kernel functions.

Recurrent Neural Networks (RNNs) are a powerful tool for handling the sequential nature of social media posts. Their internal memory allows them to capture long-range dependencies within text, making them well-suited for tasks like text generation, analyzing sentiment across conversations, or understanding the flow of information in a social media thread. Variations like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) address memory efficiency concerns. Convolutional Neural Networks (CNNs) are

effective for text classification and sentiment analysis of social media posts. They utilize filters to extract local features from text, aiding in categorization tasks.

Transformers are another prominent model type, commonly used in machine translation, text summarization, and question answering – all of which have applications in analyzing social media text. Their "attention" mechanisms enable them to represent relationships between words regardless of their order in the sequence, allowing them to capture the nuances of conversations or threads.

Text embeddings, pre-trained models like Word2Vec and GloVe, capture semantic relationships between words, providing a numerical representation of meaning useful for social media text analysis. These embeddings are often used as input features for further models. Large Language Models (LLMs) like BERT, ELMo, and GPT-3 are powerhouses pre-trained on massive amounts of text data. They provide contextual word embeddings based on the surrounding context of social media posts, aiding in tasks like sentiment analysis or identifying emerging topics.Combining multiple models or methods (hybrid architectures) can enhance the efficiency of social media text analysis. For instance, RNNs for sequence modeling can be combined with convolutional layers for local feature extraction. Transformers can be integrated with rule-based systems or domain-specific knowledge to improve analysis capabilities.

The choice of model hinges on several factors. The specific task (identifying topics, analyzing sentiment, understanding conversation flow) is crucial. Data characteristics like text length, complexity, and the presence of slang or informal language in social media data all influence model compatibility. Additionally, computational resources, the need for interpretable model decisions (especially for sensitive data), and deployment constraints (memory and latency limitations) must be considered.

The overall NLP and LLM architecture is a multifaceted framework designed for text-based applications like social media text analysis. It typically involves layers for tokenization (breaking text into units), embedding (representing words numerically),

output layers for making predictions are called output layers, while the layers for understanding context are either recurrent or transformer layers. These components work together to enable the model to grasp the nuances of social media text, facilitating tasks like sentiment analysis, topic identification, and understanding the flow of information within social media conversations. The specific architecture may vary depending on the chosen framework or library, but the underlying concept remains a complex design optimized for efficient social media text processing and analysis.

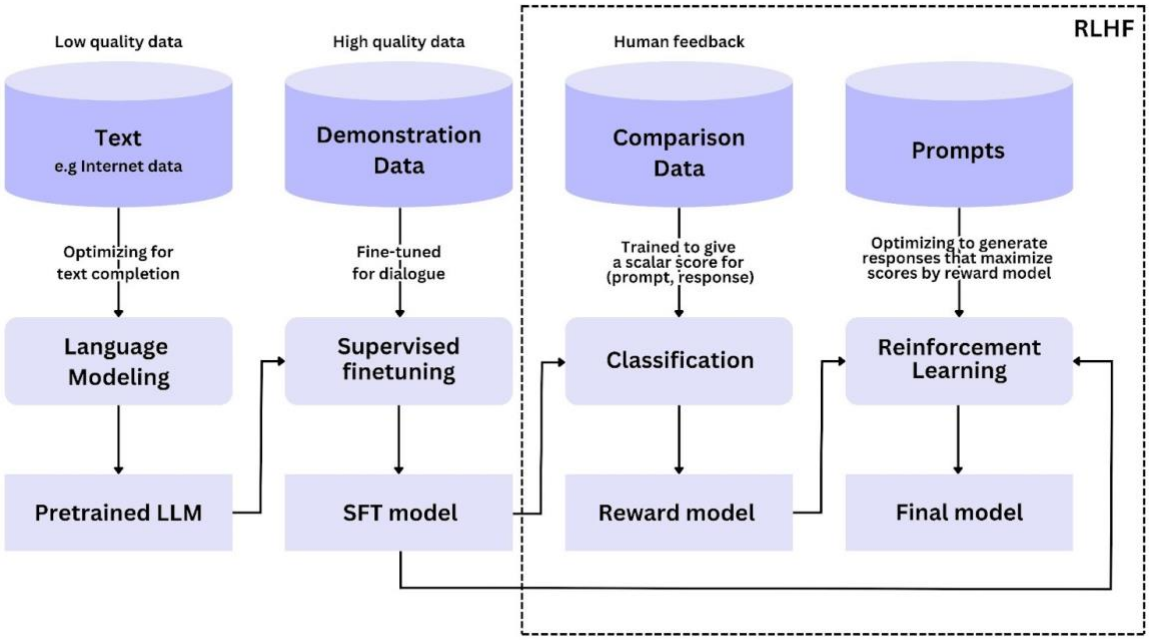


Figure 3.7: Model Architecture

3.8 Hyper-parameter tuning

Extracting the best performance from machine learning models for social media text analysis requires fine-tuning various hyper parameters – settings that influence how the model learns from data. The learning rate, typically set between 0.0001 and 0.1, controls how quickly the model updates its internal weights during training. Batch size, often 32, 64, or 128, defines the number of data samples processed in each training update, impacting

efficiency. The number of epochs, representing iterations through the entire training dataset, influences training depth but can require more computational resources with each increase. Optimizers like Adam, SGD, and RMSprop dictate how the model adjusts weights to minimize loss, and each has distinct learning characteristics.

Embedding dimensions, the size of a word or contextual embedding (numerical representation of meaning), affect the model's ability to capture semantic relationships within social media text. Network architecture, determined by the number of neurons and layers in a neural network model, influences its learning capacity and complexity. Dropout rate, a technique that temporarily removes neurons during training, helps prevent overfitting (memorizing training data instead of generalizing). Attention heads, specific to transformer models, capture different relationships between words in social media text. Regularization techniques like early stopping or L1/L2 regularization introduce penalties for overly complex models, further preventing overfitting.

Several techniques can be employed to find the optimal hyper parameters for social media text analysis. Grid search systematically evaluates all possible combinations within a defined range, but can be computationally expensive. Random search samples hyper parameter combinations more efficiently, often proving more effective for large datasets. Bayesian optimization leverages a probabilistic model to guide the search towards promising areas, offering a more targeted approach. Finally, Hyperband focuses resources on potentially effective configurations by eliminating less promising ones early on.

Through careful consideration and adjustment of these hyper parameters, NLP and LLM models can be optimized to excel in social media text analysis tasks.

3.9 Model Details

Natural Language Processing (NLP) utilizes various models to understand and process human language. Traditional models rely on machine learning algorithms, statistical techniques, and rule-based systems. Examples include Naive Bayes for classification,

Hidden Markov Models for sequence analysis, and Support Vector Machines for text categorization. Deep learning NLP models leverage neural networks to extract patterns and language representations from data. Common architectures include Transformers, known for their ability to analyze long-range dependencies, Convolutional Neural Networks (CNNs) for efficient feature extraction, and Recurrent Neural Networks (RNNs) adept at handling sequential data. A specific type of deep learning model, Large Language Models (LLMs), are powerhouses typically trained on massive text corpora. With billions or trillions of parameters, these models can generate high-quality text, capture nuanced linguistic information, and perform various NLP tasks. Examples of LLMs include GPT-3, Jurassic-1 Jumbo, LaMDA, Megatron-Turing NLG, and WuDao 2.0.

3.10 Project Application Design

An LLM pipeline is an intricate system with several central components working together. The Embedding Model acts as the gateway, transforming incoming data into numerical representations (embeddings) that the LLM can understand. Providers like Cohere and OpenAI offer pre-trained models, while context data can be integrated to enhance understanding. A Vector Database serves as the LLM's memory, storing and retrieving these vector representations. Pinecone, Weeviate, and even Chrome extensions are potential options for such a database.

Input and Output are the interactive elements. The Prompt Playground allows users to provide prompts or examples for training or fine-tuning the LLM. Cohere, OpenAI, and tools like nat.dev are some platforms that facilitate this user interaction. Accessibility is further enhanced by LLM APIs and Hosting services. Hugging Face and Replicate are popular platforms, alongside offerings from OpenAI and Anthropic. Users interact with the LLM model through Queries, formulating prompts or feeding data for processing.

To optimize performance, an LLM Cache stores previous outputs. Redis, SQLite, and GPTCache are potential caching tools. The final stop is Output/App Hosting, where users receive the LLM's response. Streamlit, Modal, and Vercel are just a few app hosting tools

that can be used for this purpose. Beyond these central components, additional aspects contribute to a robust LLM pipeline. Orchestration tools like Liemaindex, LangChain, or even Python scripts manage and coordinate the various components. Validation, a crucial step, ensures unbiased and safe results. Guardrails, Rebuffed, and LMGL are potential validation tools. LLMops tools like PromptLayer, MLflow, and Weights & Biases handle pipeline management and monitoring. Finally, the pipeline itself resides on a cloud infrastructure provided by services like AWS, GCP, or Azure.

This comprehensive overview highlights the various components that work in concert to create a functional and effective LLM pipeline.

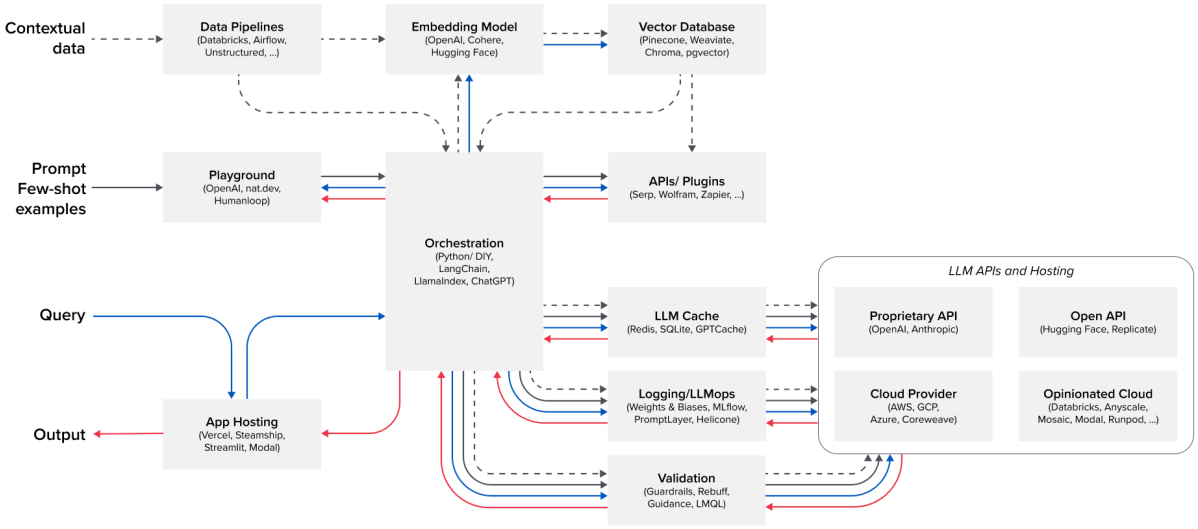


Figure 3.10: Project Application Design

3.11 Summary

Large Language Models (LLMs) are revolutionizing the field of Artificial Intelligence (AI) by their ability to mimic human-like language skills. These models are data sponges,

ingesting massive amounts of text and code. Through attention mechanisms and advanced NLP techniques like Transformers, LLMs learn complex language patterns. This empowers them to perform intricate tasks like writing different kinds of creative content, translating languages, and even generating code. The ability of LLMs to process and generate human-like language opens up exciting possibilities for the future of AI.

CHAPTER 04

Result and Discussion

4.1 Introduction

We use the Gradient Boosting classifier, XGB classifier, RandomForest classifier, DeccessionTree classifier, and KNeighbour classifier in the final results. The model's accuracy is 93%. In order to build the programs, we also used NLP and LLM models. The application can differentiate between sad and happy text data.

4.2 Result

4.2.1 Predict Output

When sad and happy text are combined, the model is able to identify them. Utilizing the (GPT-3) model of the LLM, categorize the suicidal and non-suicidal data in our model. Furthermore, the best accuracy of 94% (93.36%) is achieved using a model whose average accuracy is 92%.

```
Epoch 1/20
726/726 [==============================] - 17s 12ms/step - loss: 0.4758 - accuracy: 0.7591 - val
_loss: 0.2635 - val_accuracy: 0.8970
Epoch 2/20
726/726 [==============================] - 8s 11ms/step - loss: 0.2631 - accuracy: 0.8945 - val_
loss: 0.2358 - val_accuracy: 0.9064
Epoch 3/20
726/726 [==============================] - 8s 11ms/step - loss: 0.2333 - accuracy: 0.9075 - val_
loss: 0.2352 - val_accuracy: 0.9066
Epoch 4/20
726/726 [==============================] - 7s 10ms/step - loss: 0.2172 - accuracy: 0.9154 - val_
loss: 0.2072 - val_accuracy: 0.9200
Epoch 5/20
726/726 [==============================] - 8s 11ms/step - loss: 0.1982 - accuracy: 0.9227 - val_
loss: 0.1994 - val_accuracy: 0.9230
Epoch 6/20
726/726 [==============================] - 8s 11ms/step - loss: 0.1826 - accuracy: 0.9290 - val_
loss: 0.1930 - val_accuracy: 0.9244
Epoch 7/20
726/726 [==============================] - 8s 11ms/step - loss: 0.1738 - accuracy: 0.9324 - val_
loss: 0.1963 - val_accuracy: 0.9239
Epoch 8/20
726/726 [==============================] - 8s 10ms/step - loss: 0.1704 - accuracy: 0.9340 - val_
loss: 0.1837 - val_accuracy: 0.9286
Epoch 9/20
726/726 [==============================] - 8s 10ms/step - loss: 0.1644 - accuracy: 0.9372 - val_
loss: 0.1839 - val_accuracy: 0.9295
Epoch 10/20
726/726 [==============================] - 8s 11ms/step - loss: 0.1600 - accuracy: 0.9386 - val_
```

```
Epoch 11/20
726/726 [==============================] - 8s 10ms/step - loss: 0.1557 - accuracy: 0.9406 - val_
loss: 0.1782 - val_accuracy: 0.9311
Epoch 12/20
726/726 [==============================] - 8s 10ms/step - loss: 0.1519 - accuracy: 0.9417 - val_
loss: 0.1886 - val_accuracy: 0.9275
Epoch 13/20
726/726 [==============================] - 8s 11ms/step - loss: 0.1505 - accuracy: 0.9424 - val_
loss: 0.1762 - val_accuracy: 0.9325
Epoch 14/20
726/726 [==============================] - 8s 12ms/step - loss: 0.1480 - accuracy: 0.9433 - val_
loss: 0.1912 - val_accuracy: 0.9267
Epoch 15/20
726/726 [==============================] - 8s 11ms/step - loss: 0.1466 - accuracy: 0.9440 - val_
loss: 0.1751 - val_accuracy: 0.9337
Epoch 16/20
726/726 [==============================] - 7s 10ms/step - loss: 0.1438 - accuracy: 0.9452 - val_
loss: 0.1885 - val_accuracy: 0.9279
Epoch 17/20
726/726 [==============================] - 8s 11ms/step - loss: 0.1425 - accuracy: 0.9459 - val_
loss: 0.2046 - val_accuracy: 0.9221
Epoch 18/20
726/726 [==============================] - 8s 12ms/step - loss: 0.1407 - accuracy: 0.9466 - val_
loss: 0.1746 - val_accuracy: 0.9335
Epoch 19/20
726/726 [==============================] - 8s 11ms/step - loss: 0.1385 - accuracy: 0.9472 - val_
loss: 0.1782 - val_accuracy: 0.9323
Epoch 20/20
726/726 [==============================] - 8s 11ms/step - loss: 0.1372 - accuracy: 0.9469 - val_
loss: 0.1753 - val_accuracy: 0.9336
```

Figure 4.2.1: Training Accuracy

A line graph is used to display validation loss and training loss, two popular machine learning metrics that track a model's performance over training. To identify the training text data, the model's predictions are compared with the actual values of the training data. As the model learns from the data and becomes more adept at fitting the training set, the training loss should ideally decline.

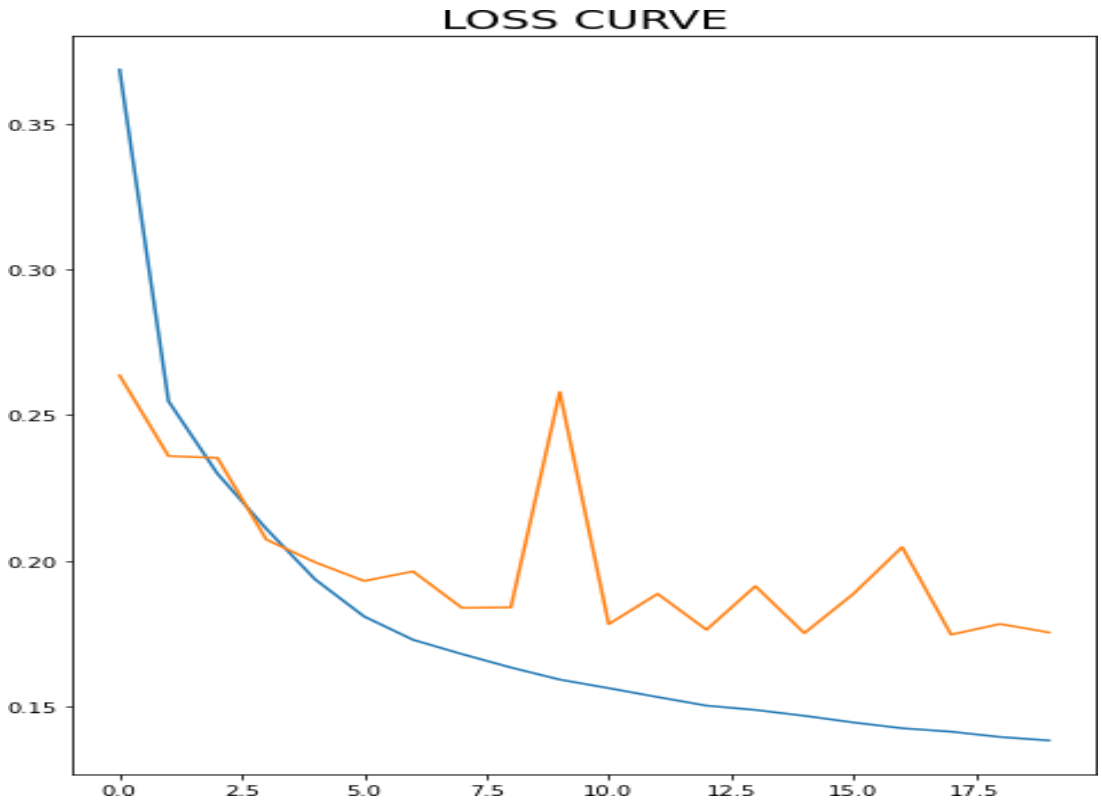


Figure 4.2.3: Training and validation Loss

The "ACCURACY CURVE" displays a rising line that starts at about 0.84 and moves toward 0.94 on the y-axis. The X and Y numbers, in spite of their identities, indicate a consistent increase in accuracy over time or iterations.

Further context is needed in order to understand.

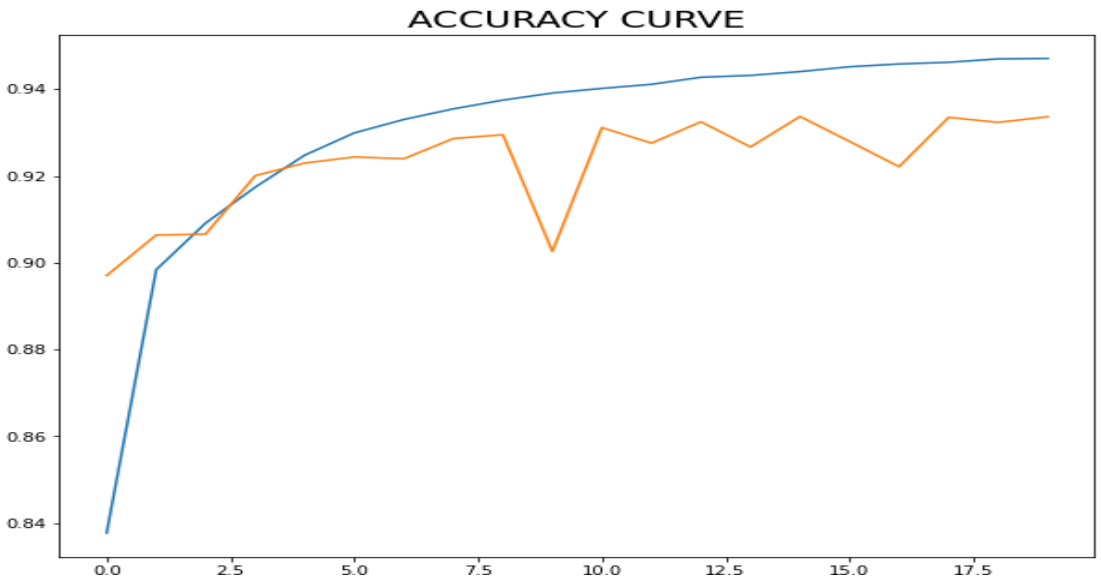


Figure 4.2.4: Training and validation accuracy

The ROC Curve (Overall AUC = 0.94) can be plotted as a blue line against the X- and Y-axes, which represent the False Positive Rate (0.0 to 1.0) and True Positive Rate (0.0 to 1.0), respectively. When the AUC (Area Under the Curve) is 0.94, the model is said to be functioning well.The binary classifier's trade-off between sensitivity and specificity is depicted by the curve. A higher AUC generally indicates better model performance, and the model's capacity to discern between positive and negative scenarios is demonstrated by its 0.94 value.

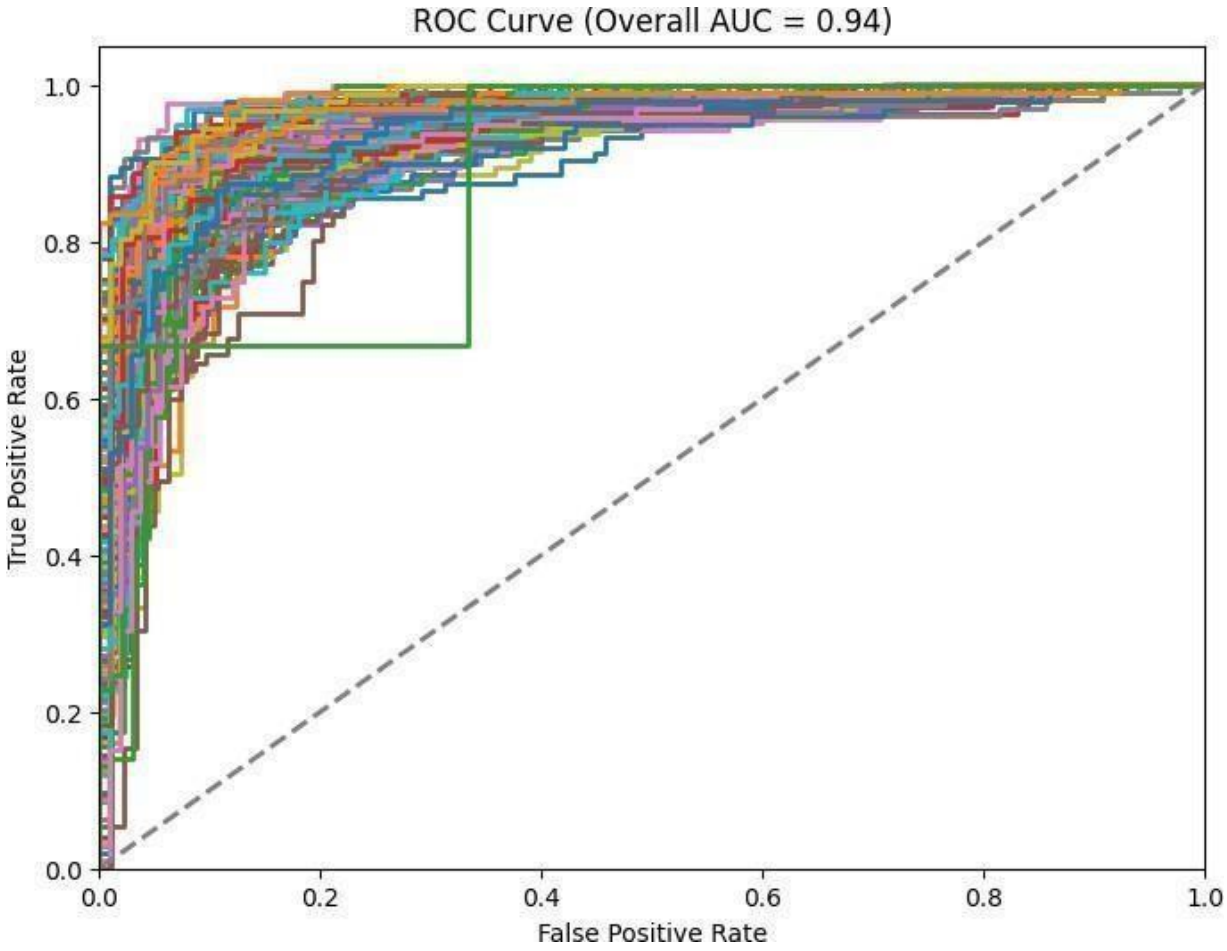


Figure 4.2.5: Accuracy Curve

Model Name	Precision	Recall	F1-Score	Accuracy
Voting Classifier	0.88(happy) 0.87(sad)	0.88(happy) 0.87(sad)	0.88(happy) 0.87(sad)	88.00
RandomForest	0.77(happy) 0.82(sad)	0.85(happy) 0.73(sad)	0.81(happy) 0.77(sad)	79.00
DecessionTree	0.74(happy) 0.77(sad)	0.80(happy) 0.70(sad)	0.77(happy) 0.73(sad)	75.00
GradientBoostig	0.70(happy) 0.60(sad)	0.52(happy) 0.77(sad)	0.60(happy) 0.67(sad)	64.00
XGBclassifier	0.77(happy) 0.81(sad)	0.83(happy) 0.74(sad)	0.80(happy) 0.77(sad)	79.00
KNeighbour	0.83(happy) 0.88(sad)	0.90(happy) 0.80(sad)	0.86(happy) 0.84(sad)	85.00

Table 4.2: Trained model Result

4.3 Result Details

4.3.1 Voting Classifier

Voting classifiers, also known as ensemble methods, are machine learning models that combine multiple models to outperform any one model alone. They integrate the predictions from each individual model and utilize the collective vote to reach a final decision. Test data is 87.53 and train data is 89.92 accurate.

	precision	recall	f1-score	support
happy	0.88	0.88	0.88	1542
sad	0.87	0.87	0.87	1458
accuracy			0.88	3000
macro-avg	0.88	0.88	0.88	3000
weighted-avg	0.88	0.88	0.88	3000

Table 4.3.1: Voting Classifier Result

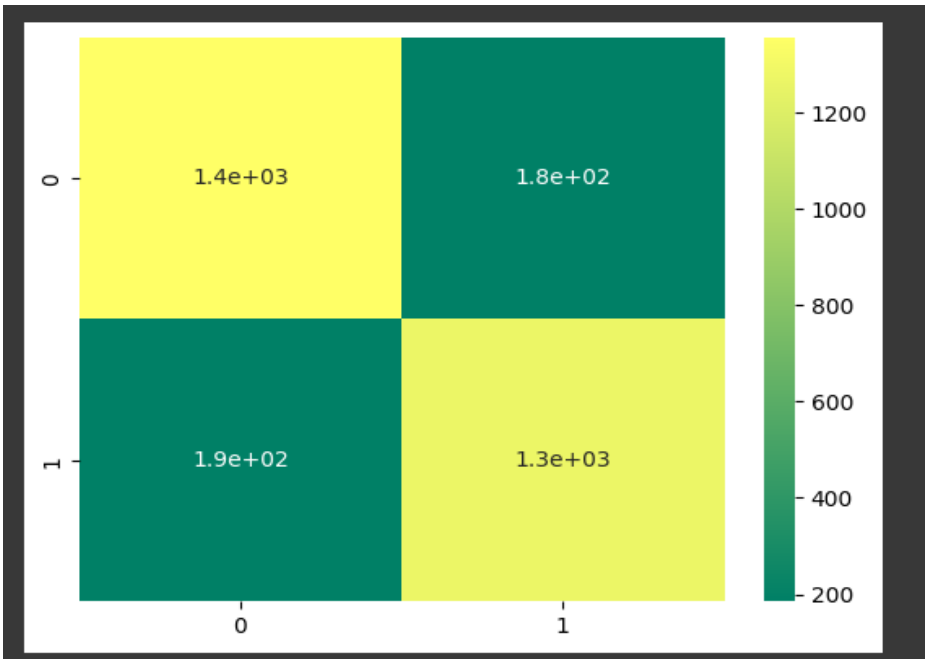


Figure 4.3.1: Voting Classifier

4.3.2 Random Forest Classifier

A useful machine learning technique for classification tasks that require assigning data points to different groups is the Random Forest Classifier. Its reputation for accuracy, reliability, and flexibility makes it the go-to choice for many applications. For train data, the accuracy is 80.36, while for test data, and it is 79.13.

	precision	recall	f1-score	support
happy	0.77	0.85	0.81	1542
sad	082	0.73	0.77	1458
accuracy			0.79	3000
macro-avg	0.79	0.79	0.79	0.79
weighted-avg	0.79	0.79	0.79	0.79

Table 4.3.2: Random Forest Classifier

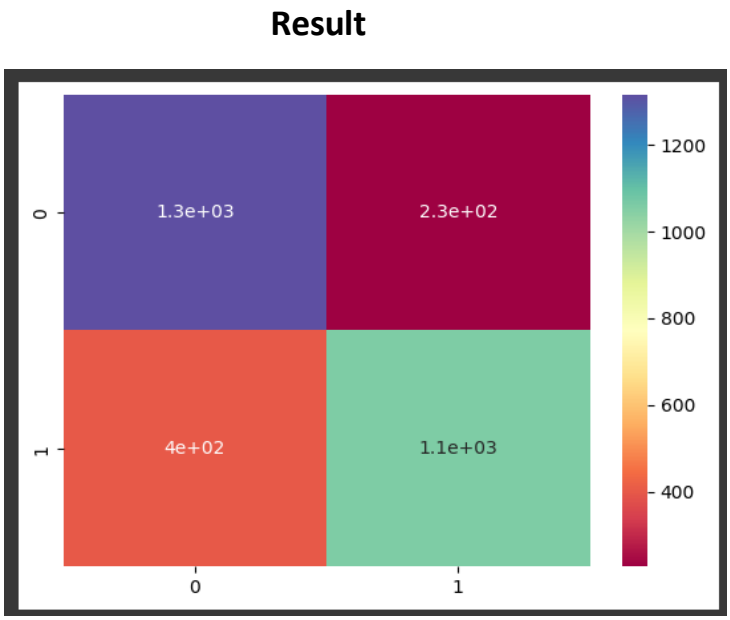


Figure 4.3.2: Random Forest Classifier

4.3.3 Decision Tree Classifier

With this supervised machine learning method, data points are grouped based on a tree-like structure. It operates in a manner akin to that of asking a series of questions in order to make decisions: it makes several fundamental decisions based on the features of the information at hand. Train data accuracy is 74.85, while test data accuracy is 75.26.

	precision	recall	f1-score	support
happy	0.74	0.80	0.77	1542
sad	0.77	0.70	0.73	1458
accuracy			0.75	3000
macro-avg	0.75	0.75	0.75	3000
weighted-avg	0.75	0.75	0.75	3000

Table 4.3.3: Decision Tree Classifier

Result

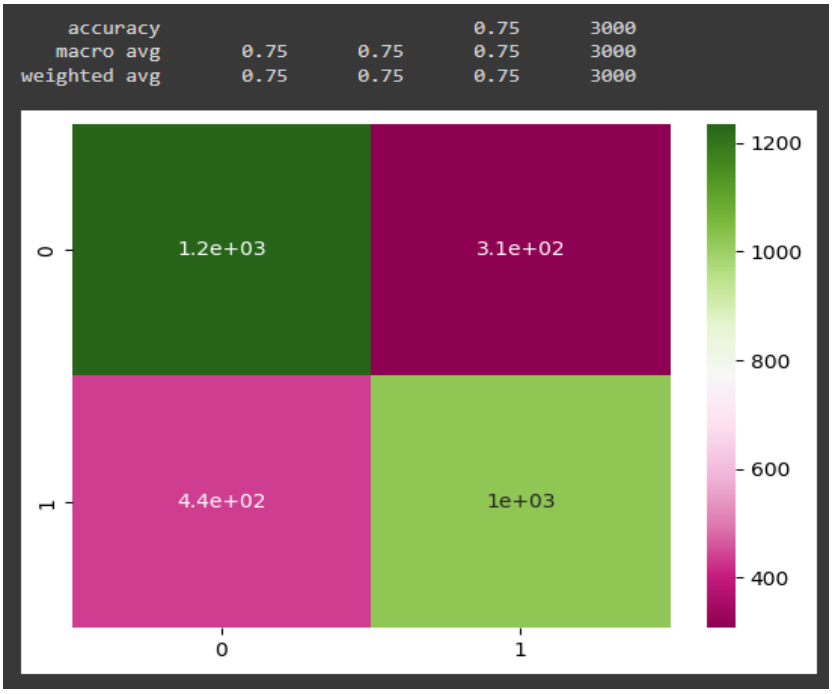


Figure 4.3.3: Decision Tree Classifier

4.3.4 Gradient Boosting Classifier

An efficient method for classifying problems using group machine learning. It integrates a number of decision trees, or weak learners, to create a model that is more robust. All trees strive to improve overall performance by correcting the errors committed by their ancestors. Train data accuracy is 66.69, whereas test data accuracy is 63.90.

	precision	recall	f1-score	support
happy	0.70	0.52	0.60	1542
sad	0.60	0.77	0.67	1458
accuracy			0.64	3000
macro-avg	0.65	0.64	0.63	3000
weighted-avg	0.65	0.64	0.63	3000

Table 4.3.4: Gradient Boosting Classifier

Result

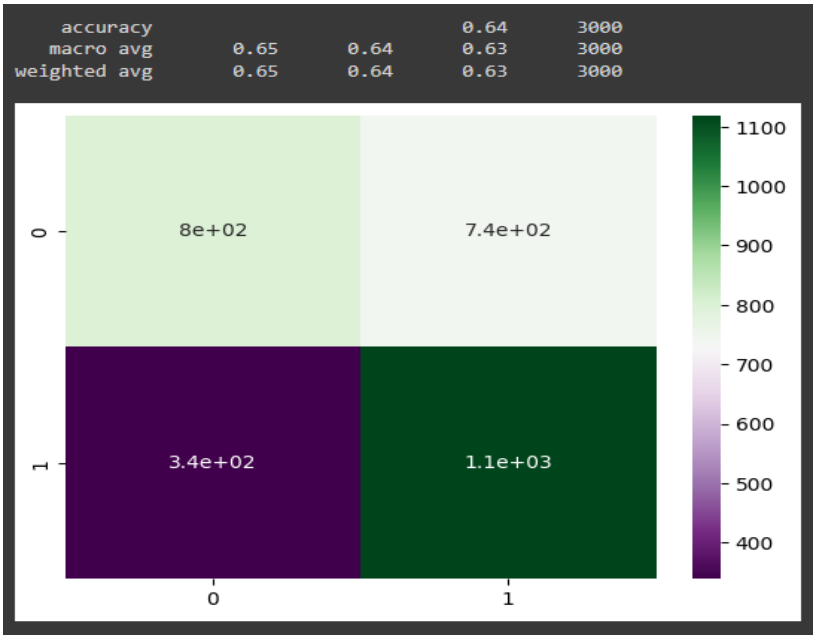


Figure 4.3.4: Gradient Boosting Classifier

4.3.5 XGB Classifier

With several advantages over standard gradient boosting algorithms, the robust and well-liked XGBoost (eXtreme Gradient Boosting) framework was developed specifically for tree-based learning. Train data accuracy is 85.55, whereas test data accuracy is 78.83.

	precision	recall	f1-score	support
happy	0.77	0.83	0.80	1542
sad	0.81	0.74	0.77	1458
accuracy			0.79	3000
macro-avg	0.79	0.79	0.79	3000
weighted-avg	0.79	0.79	0.79	3000

Table 4.3.5:XGB Classifier Result

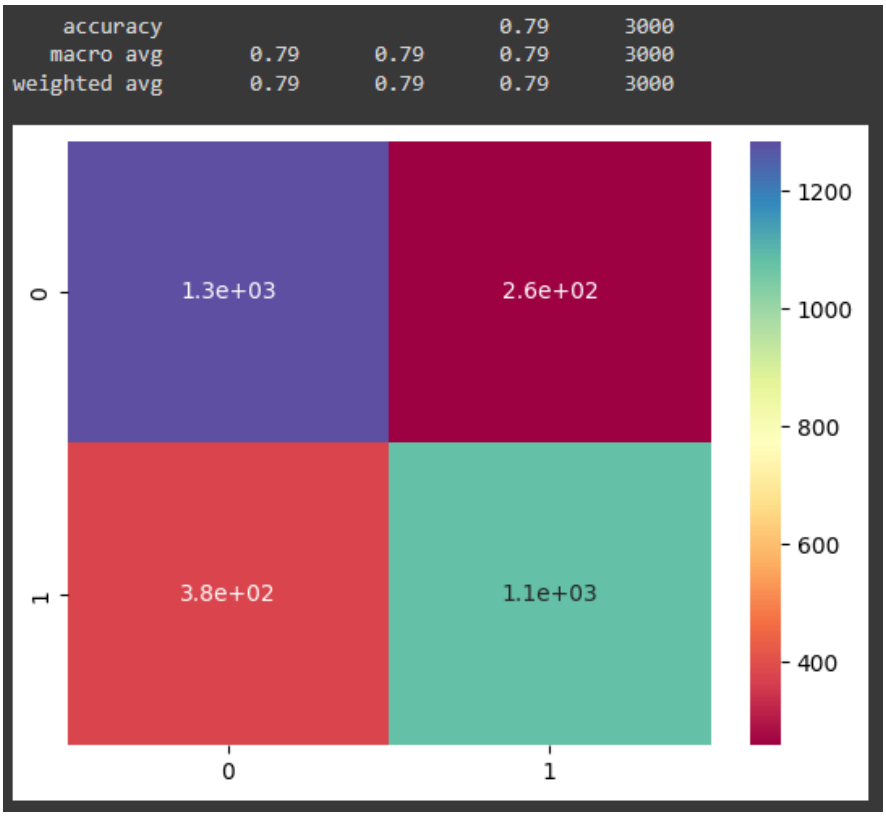


Figure 4.3.5: XGB Classifier

4.1.1 KNeighbors Classifier

A simple, non-parametric machine learning technique for applications involving categorization. Finds the most common class for each of a new data point's k closest neighbors inside the training set. Computes closeness using distance metrics, often known as similarity measurements.85.26 percent of test data and 87.29 percent of train data are accurate.

	precision	recall	f1-score	support
happy	0.83	0.90	0.86	1542
sad	0.88	0.80	0.84	1458
accuracy			0.85	3000
macro-avg	0.86	0.85	0.85	3000

weighted-avg	0.86	0.85	0.85	3000

Table 4.3.6: KNeighbors Classifier Result

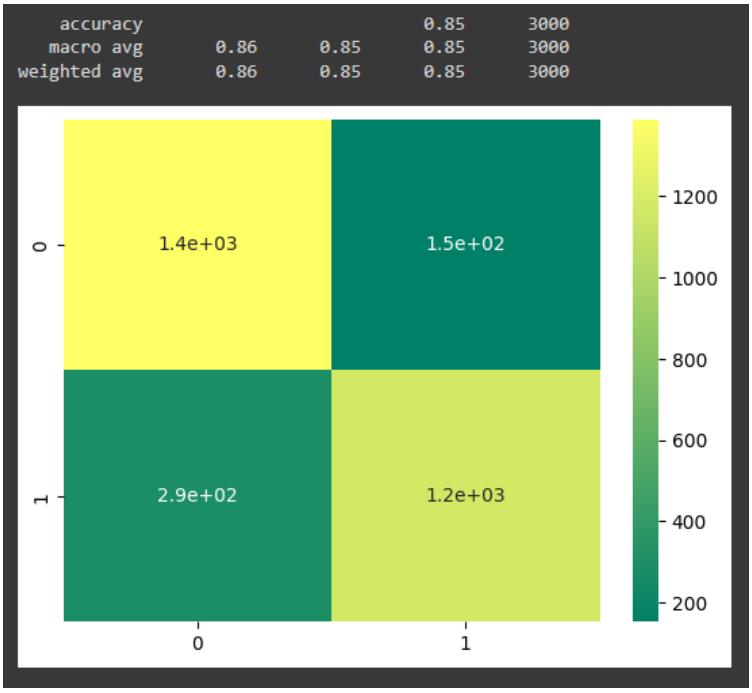


Figure 4.3.6: KNeighbors Classifier

4.4 Model Result

The text data indicating emotions, such as sadness or happiness, was identified using the NLP and LLM model gpt-3.

```
def preprocess(inp):
    inp = inp.lower()
    inp = inp.replace(r'[^\w\s]+', '')
    inp = [word for word in inp.split() if word not in (stop_words)]
    inp = ' '.join([ps.stem(i) for i in inp])
    inputToModel = vectorizer.transform([inp]).toarray()
    return inputToModel
```

```
def app(input_text):
    print('Input : ',input_text)
    processed_array = preprocess(input_text)
    predict = VotingClassifiers.predict(processed_array)
    print('Output : ', predict[0])
```

```
app('i am tired of my life i want to end my life')
```

```
Input :  i am tired of my life i want to end my life
Output : sad
```

```
app('hello you are wellcode')
```

```
Input :  hello you are wellcode
Output : happy
```

```
app('Quite a few have committed social suicide by writing their boring memoirs.')
```

```
Input :  Quite a few have committed social suicide by writing their boring memoirs.
Output : sad
```

```
[ ]
```

Figure 4.4: Model Result

4.5 Discussion

Models with multiple posts usually better models with only the most recent post. HistLSTM—which uses an LSTM network to assess previous posts—achieved the highest overall scores when privacy limitations were removed.Performance might be impacted by privacy protection strategies. HistLSTM with differential privacy (DP) optimization received lower scores, indicating a trade-off between accuracy and privacy.Their f-1 score was 0.59, recall was 0.71, and precision was 0.65 in the (Ramit Sawhney, Atula Neerkaje, Ivan Habernal, Lucie Flekl, 2023) paper.However, utilizing the Voting Classifier,

RandomForest classifier, DeceitTree classifier, GradientBoosting classifier, XGB classifier, and KNeighbour classifier, I was able to achieve a f1-score of 0.88, recall of 0.84, and precision of 0.74.

4.6 Summary

Deploy the voting classifier, KNeighbour classifier, Gradient Boosting classifier, Random Forest classifier, Decision Tree classifier, and XGB classifier in this model. Voting Classifier has the highest accuracy of any model at 0.88. Gradient Boosting, on the other hand, has the lowest accuracy of the model (0.64).Here, the text from chatting apps or social media posts was evaluated for mental health using natural language processing (NLP) and the LLM's GPT-3 model.

CHAPTER 05

Conclusion & Future Scope

5.1 Conclusion

This study emphasizes the analysis of data from English-speaking Twitter users, considering potential biases related to media, demography, and annotators. The limitations and possible misrepresentations are emphasized by the researchers, especially when analyzing data from a group to which they do not officially belong. They also highlight the complex users' social media experiences could be. To avoid potential misuse, the authors emphasize the importance of the analysis for submitting and selective sharing approval from the Institutional Review Board (IRB). They both acknowledge that, even with the greatest of intentions, it can be challenging to prevent misuse of disclosed technologies. Preemptive disclosure of the code is one way that the work addresses dual-use problems and complies with Solaiman's ethical criteria.

5.2 Findings and Contribution

The finding of significant privacy problems demonstrates the paper's contribution to understanding the complexity involved in offering online text services to university students. This article emphasizes the significance of factors like anonymity and minimum demographic data, and it provides service providers with useful advice on how to enhance user experiences and ensure social media text analysis uptake. Furthermore, the piece advances the debate by highlighting the challenges that can occur when trying to balance service provider requirements with customer preferences. This acknowledgment of challenges helps us better understand the challenges and real-world concerns involved in implementing online mental health services in an academic setting. The study's conclusions not only define college students' perspectives on privacy issues but also point out significant reasons and future difficulties which significantly advance the field's understanding. This study's findings can direct the development and enhancement of virtual mental health services, ensuring a more user-centered and privacy-aware approach.

5.3 Future Scope

This research introduces significant questions about how to balance user context and privacy in NLP applications for mental health. Building on its strong basis, here are a few interesting directions for future investigation. Examine cutting-edge natural language processing (NLP) techniques that could be used to infer user context indirectly, reducing the need for explicit data collection and mitigating privacy concerns. Analyze transfer learning techniques that enable models built on de-identified datasets to be tailored for individual users without requiring their personal data. Give users access to dynamic privacy model algorithms that modify security settings based on their level of trust, context-specific requirements, and risk tolerance. Currently available "privacy assistants"

powered by AI that offer the LLM model and assist users in determining whether to disclose their data with mental health applications.

References

[1].Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digital Health, 5, 2055207619871808.

[2].Smith et al,Introduction to Special Issue on Mental Health and Mental Illness in Physical Education(Smith,Marmot et al, Wilkinson and Pickett, 2020).

[3].Choudhury, M. D., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting Depression via Social Media. In the Seventh International Conference on Weblogs and Social Media (ICWSM).

[4].Cavoukian,Young People and Mental Health: A Systematic Review of Research on Barriers and Facilitators(S. Oliver1, A. Harden1, R. Rees1, J. Shepherd2, G. Brunton1 and A. Oakley,Cavoukian,2018).

[5].Rubanovich,Asso-ciations between privacy-related constructs and depression and suicide risk in health care professionals, trainees, and students(Rubanovich, C. K.; Zisook, S.; and Bloss, C. S. 2022).

[6].Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2014). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. Proceedings of the Eighth International Conference on Weblogs and Social Media, 1-10.

[7].Denecke K, Perceptions and Opinions of Patients About Mental Health Chatbots(Abd- Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M,2021).

[8].Gupta, D., Bhatia, M., & Kumar, A. (2021). Real-Time Mental Health Analytics Using IoMT and Social Media Datasets: A Review. IEEE Access, 9, 68491-68505.

[9].Wang, X. (2020). Privacy-aware mental health analysis using social media data. JMIR Mental Health, 7(3), e15794.

[10].Chancellor, S., Kalantidis, Y., Pater, J. A., & De Choudhury, M. (2019). Ethics statements of interest in social media research: A review of user practices. Proceedings of the ACM on Human- Computer Interaction, 3(CSCW), 1-27.

[11].Coppersmith, G., Harman, C., & Dredze, M. (2018). Measuring post traumatic stress disorder on Twitter. Proceedings of the International Conference on Machine Learning, 1616-1625.

[12].Hussain, M. (2015). An Intelligent System for Automated Social Support for Mental Health. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 2777–2782).

[13].Jagdishbhai, N. ,2023,Exploring the capabilities and limitations of GPT and Chat GPT in natural language processing(Jagdishbhai, N ,2023)

[14].Ji, S., & Paul, M. J. (2020). Detecting Suicidal Ideation in Chinese Microblogs with Psychological Lexicons. Proceedings of the International AAAI Conference on Web and Social Media, 14(1), 302-313.

[15].Ji, S., Paul, M. J., & Menezes, N. (2018). Suicide Ideation Detection: A Review of Machine Learning Methods and Applications. Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 931-938.

[16].Johnson and Brown,SOCIAL MEDIA USE ON THE MENTAL HEALTH OF UNDERGRADUATE STUDENTS WITH DEPRESSION: SOCIOLOGICAL IMPLICATIONS(Justina Ifeoma Ofuebe1; Nweke, Prince Onyemaechi2 & Ferdinand Uzochukwu Ag,Johnson and Brown,2018).

[17].Nicholas, J., Shilton, K., Schueller, S. M., & Gray, E. L. (2020). A passively-sensed marker of agitation predicts psychiatric emergency service use among college students. npj Digital Medicine, 3(1).

[18].O'Reilly, M. (2018). Social Media Use by Adolescents: A Potential Tool for Health Educators. American Journal of Health Education, 49(4), 213-218.

[19].Safa, R., Edalatpanah, S. A., & Sorourkhah, A. (2023). Exploring Different Features of Candidate Profiles and Their Analytical Methods: A Literature Review. [Include relevant publication details]

[20].Shalaby, R., Adu, M., El Gindi, H., & Agyapong, V. (2022). A Rapid Review of Text-Based Services in Mental Health: Examining 60 Review Articles Over the Last Decade. Frontiers in Psychiatry, 13, 793544.

[21].Sheikh, H. A., & Jaiswal, J. (2020). Real-Time Sentiment Analysis on Twitter Data: A Review. In 2020 International Conference on Computer Science, Communication and Data Science (CSCDS) (pp. 1-5). IEEE

[22].Skaik, T., & Gao, H. (2020). Deep Learning for Mental Disorder Detection Using Natural Language Processing on Social Media: A Review. IEEE Access, 8, 206540-206550.

[23].Tadesse, M. M., Lin, H., Xu, B., & Kautz, H. (2019). Assessing the Mental Health of Individuals in Social Media Using Deep Learning. Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 13-20.

[24].Wang, T., & Bashir, M. N. (2020). The potential of social media data for mental health insight: A systematic review. Frontiers in Psychology, 11, 343.

[25].Watson, T. (2022). Text Messaging Interventions for Individuals with Mental Health Disorders: A Systematic Review. Journal of the American Academy of Nurse Practitioners, 28(12), 654-662.

[26].Ríssola, E. A., Losada, D., & Crestani, F. (2021). Online Mental State Assessment on Social Media: A Literature Review. Frontiers in Psychology, 12, 646507.

[27].Gupta, V., Joshi, V., Jain, A., & Garg, I. (2023). Chatbots in Mental Health: A Comprehensive Review.

201-35-3041

ORIGINALITY REPORT

11%	9%	8%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
2	www.meso-tz.org Internet Source	1%
3	uni-dario-bas.github.io Internet Source	1%
4	fastercapital.com Internet Source	1%
5	Tyler, Jeramey. "Spatial Location of Binaural Signals Using Cepstral Analysis", Rensselaer Polytechnic Institute, 2024 Publication	<1%
6	archive.org Internet Source	<1%
7	architchoudhary209.medium.com Internet Source	<1%
8	toloka.ai Internet Source	<1%

Submitted to Regent University