Leveraging Machine Learning for Depression Identification in Social Media

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Technology

In

Computer Science and Engineering

Joy Aloysius 20BCE2337

Vishvesh Mishra 20BCE2362

Kartikey Kesarwani 20BCE2373

Under the guidance of

Dr. Poornima N,

Assistant Professor Senior Grade 1,

SCOPE, VIT, Vellore

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

VELLORE INSTITUTE OF TECHNOLOGY

VELLORE 632 014, TAMIL NADU, INDIA

May 2024

DECLARATION

I hereby declare that the thesis entitled "Leveraging Machine Learning for Depression Identification in Social Media" submitted by me, for the award of the degree of Bachelor of Technology in Computer Science and Engineering to VIT is a record of bonafide work carried out by me under the supervision of Prof. / Dr. Poornima N.

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place : Vellore

Date :

Signature of the Candidate

CERTIFICATE

This is to certify that the thesis entitled "Leveraging Machine Learning for Depression Identification in Social Media" submitted by *Kartikey Kesarwani 20BCE2373,Vishvesh Mishra 20BCE2373, Joy Aloysius 20BCE2337* **School of Computer Science and Engineering**, VIT, for the award of the degree of Bachelor of technology in Computer Science and Engineering, is a record of bonafide work carried out by him/her under my supervision during the period, 01. 12. 2023 to 30.04.2024, as per the VIT code of academic and research ethics.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The thesis fulfills the requirements and regulations of the University and in my opinion meets the necessary standards for submission.

Place: Vellore

Date :

Signature of the Guide

Internal Examiner

External Examiner

Dr. UMADEVI KS

Computer Science and Engineering

ACKNOWLEDGEMENTS

We would like to extend my heartfelt thanks to my guide, Prof. Poornima N, for her invaluable guidance, support, and mentorship throughout the process of writing this thesis. Her extensive knowledge and insightful feedback have been instrumental in shaping the direction and quality of this work. We deeply appreciate his dedication and patience, which have greatly contributed to my academic growth and success. In addition, we would like to express our sincere appreciation to VIT for providing us with the resources and opportunities necessary for our academic journey. The supportive academic environment, exceptional faculty, and access to state-of-the-art facilities and research materials have been critical in enabling us to complete this thesis. We are particularly thankful for the collaborative atmosphere at VIT, which has fostered intellectual curiosity and encouraged us to pursue this research topic. Without the support and guidance of Prof. Poornima N and the resources provided by VIT, this thesis would not have been possible. We are deeply indebted to everyone who has played a role in this endeavor.

Thank you.

Student Name

Executive Summary

Our project focuses on developing a robust system for detecting signs of depression in social media posts, particularly on platforms like Twitter and Instagram. Leveraging advanced deep learning techniques, including Bidirectional LSTM and Multi-head Attention mechanisms, our system processes unstructured text data to identify subtle indicators of depression. Additionally, we integrate a text extraction from image module to extract text from images containing social media posts, further enhancing the scope of our analysis.

The core of our model lies in its ability to understand the context and sentiment of social media posts, enabling it to accurately classify them as depressive or non-depressive. Furthermore, we incorporate the GPT API to generate personalized messages for users identified as potentially experiencing depression, providing empathy, encouragement, and resources for seeking help. By deploying our system, we aim to contribute to the early detection and intervention of mental health issues in online communities, ultimately fostering a more supportive and compassionate digital environment. Our project holds significant implications for improving mental health outcomes by empowering individuals to seek timely support and access resources tailored to their needs. Through continuous refinement and evaluation, we strive to develop an effective tool for promoting mental well-being in the digital age.

TABLE OF CONTENTS

List of Figures

List of Abbreviations

BiLSTM	Bidirectional Long Short-Term Memory
CSV	Comma-Separated Values
API	Application Programming Interface
NaN	Not a Number
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
NN	Neural Network
BERT	Bidirectional Encoder Representations from Transformers
GPT	Generative Pre-trained Transformer
RMSprop	Root Mean Square Propagation (an optimization algorithm)
SGD	Stochastic Gradient Descent (an optimization algorithm)
Relu	Rectified Linear Unit (an activation function
ML	Machine Learning
AI	Artificial Intelligence
DL	Deep Learning
TF-IDF	Term Frequency-Inverse Document Frequency
NLP	Natural Language Processing
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
LR	Learning Rate
SGD	Stochastic Gradient Descent
MSE	Mean Squared Error
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
ROC	Receiver Operating Characteristic (curve)

TPR	True Positive Rate
FPR	False Positive Rate
AUC	Area Under the Curve
F1 Score	A metric combining precision and recall
PR Curve	Precision-Recall Curve

1. INTRODUCTION

1.1. OBJECTIVE

The objective of this research is to identify indicators of mental health issues, particularly depression, within social media content through the application of advanced natural language processing (NLP) and deep learning approaches. This paper introduces a novel model architecture designed to capture the subtleties of language, emotional cues, and social interaction patterns found on platforms like Instagram and Twitter.

Our approach incorporates a comprehensive analysis of both textual and multimedia content, recognizing that depression-related discussions often involve various forms of media, including text-based posts and images with embedded text. By utilizing cutting-edge methods to extract and process this information, our model can better understand the context and semantics associated with mental health discussions.

Furthermore, the research integrates personalized message generation, employing the GPT API to create empathetic responses. This aspect of the study aims to offer supportive resources and encouraging messages to individuals identified as potentially struggling with depression, contributing to early detection and intervention efforts.

The broader goal of this work is to promote awareness about the intersection of mental health and social media, advocating for a reduction in stigma, as well as fostering an environment that encourages early recognition and appropriate responses to mental health issues in online communities. By engaging in this research, we hope to contribute to a more compassionate and informed approach to addressing mental health challenges in the digital age.

1.2 Motivation

Our project is focused on building a comprehensive computational framework to study the link between mental health and social media. Using advanced natural language processing (NLP) and deep learning methods, we aim to identify subtle signs of mental health challenges, especially depression, in social media content. Alongside our analysis, we plan to integrate supportive features that connect algorithmic insights with a human-centered approach, promoting a more empathetic digital environment. This project tackles an important societal concern, introduces innovative solutions, and underscores the critical role of empathy in technology-driven strategies for mental health support.

1.3 Background

Mental health is a growing concern worldwide, affecting individuals from all walks of life. Depression and anxiety are among the most common mental health disorders, impacting millions of people and their families. As awareness around mental health issues increases, society is also exploring new ways to understand and address these challenges. Social media, with its widespread use and significant influence, has emerged as both a potential source of stress and a platform for support and connection. Platforms like Instagram and Twitter are central to modern communication, offering users the opportunity to share their thoughts, emotions, and daily experiences. While this increased connectivity can foster community and support networks, it also carries risks. Cyberbullying, social comparison, and the pressure to present an idealized image can contribute to negative mental health outcomes. Additionally, the addictive nature of social media, with its constant stream of content, may lead to increased anxiety and depression symptoms for some individuals.

Research into the relationship between social media use and mental health has produced mixed results. Some studies suggest that heavy social media use is associated with higher levels of depression and anxiety, possibly due to the constant comparison with others and the sense of isolation that can arise from online interactions. Conversely, other studies highlight the positive aspects of social media, such as the ability to connect with like-minded individuals and access supportive communities. These differing findings indicate a complex interplay between social media and mental health, influenced by factors like frequency of use, type of engagement, and individual personality traits.

The challenge of identifying mental health concerns in social media content lies in its inherent complexity. Language on social media is often informal, nuanced, and subject to rapid evolution. Effective analysis requires a combination of linguistic expertise and advanced computational techniques. Furthermore, ethical considerations around privacy and data protection are paramount. Researchers must balance the need for insight with the responsibility to protect users' sensitive information.

Recent advances in natural language processing (NLP) and deep learning have opened new avenues for analyzing large-scale text data. Bidirectional Long Short-Term Memory (BiLSTM) networks, with their capacity to understand context in text sequences, have shown promise in recognizing patterns that indicate mental health issues. Attention mechanisms, which allow the model to focus on the most relevant parts of the text, add an additional layer of interpretability.

Our research aims to build on these technological advancements to create a framework for detecting subtle indicators of mental health concerns in social media content. By employing BiLSTM with attention mechanisms, we can extract meaningful insights from textual data and focus on the words and phrases that signal emotional distress. This approach has the potential to help identify early signs of depression and other mental health conditions, facilitating timely interventions and support.

In integrating multimedia content analysis with textual cue detection, our study aims to bridge the gap between data-driven analysis and empathetic human understanding. The ultimate goal is to contribute to a more compassionate digital environment, where technology aids in identifying those in need of support without compromising their privacy and autonomy.

2. PROJECT DESCRIPTION AND GOALS

2.1 Survey on Existing System

Multimodal sentiment analysis sits at the intersection of various research disciplines, drawing from advancements in natural language processing (NLP), computer vision, and affective computing. Examining current systems reveals a diverse array of methodologies designed to decode the complex expressions of sentiment across multiple modalities.

The shift towards deep learning has transformed how multimodal sentiment analysis is approached. Data-driven models, with their ability to learn from large volumes of diverse data, have become the standard. Among these, Deep Multimodal Fusion Networks, Multimodal Attention Networks, and Transformer-based Models have gained traction, primarily because they excel at extracting subtle cues from various modalities. These architectures leverage the strengths of neural networks to capture complex emotional expressions that were previously elusive with traditional approaches.

Techniques such as Multimodal Attention Networks are particularly noteworthy because they allow the model to focus on the most relevant parts of input data across different modalities. This attention-based approach has been successful in identifying key features, leading to improved sentiment analysis accuracy. Similarly, Transformer-based Models, renowned for their self-attention mechanisms, have set a new benchmark for fusing distinct data types, offering a more comprehensive understanding of sentiment.

A critical component of multimodal sentiment analysis is the availability of high-quality datasets that accurately represent the diversity of sentiment in real-world scenarios. Foundational datasets like DAIC-WOZ have been invaluable in driving research on depression detection and emotional analysis. However, there's a growing need for annotated datasets that encompass a broader spectrum of emotions and contexts. This gap presents a challenge for researchers aiming to develop robust multimodal models.

The evolution of fusion techniques has played a crucial role in advancing multimodal sentiment analysis. Recent models seamlessly integrate text, audio, and visual data to provide a more nuanced interpretation of sentiment. Approaches like hierarchical fusion architectures and attention-based fusion methods have enabled more sophisticated analysis, allowing models to consider context across multiple modalities. Frameworks such as BERT-CNN for text processing and CNN-BiLSTM for audio data exemplify this trend, demonstrating that successful fusion can overcome the limitations inherent in single-modality models.

Despite the progress in multimodal sentiment analysis, several challenges remain. Ensuring consistent performance across diverse data sources, addressing data quality issues, and maintaining ethical standards are critical for advancing the field. Future research must focus on expanding datasets, refining fusion techniques, and enhancing model interpretability to build systems that can reliably analyze sentiment in complex, multimodal environments.

2.2 Research Gap

Although significant advancements have been made in multimodal sentiment analysis, there are still notable gaps that signal promising directions for further research and innovation. A critical shortcoming is the limited availability of annotated datasets that reflect a broad range of emotions and cultural contexts. Existing datasets offer valuable information on certain aspects of sentiment analysis but often lack the depth needed to represent subtle emotional states, which in turn restricts the generalizability of multimodal models.

Another persistent challenge is the interpretability and explainability of deep learning models, which can hinder the implementation of multimodal sentiment analysis in practical applications. The opaque nature of neural networks makes it difficult to understand the reasoning behind their predictions, leading to concerns and distrust among end-users and stakeholders. Addressing this challenge requires creating models that are more transparent and interpretable, providing clarity on the factors that influence sentiment predictions and building trust in these technologies.

A significant challenge in multimodal sentiment analysis is addressing the ethical and societal implications, especially regarding privacy, bias, and algorithmic fairness. The extensive use of social media data intensifies concerns about user privacy and informed consent, leading to questions about data governance and appropriate regulatory frameworks. These concerns require rigorous approaches to ensure users' rights and data protection are respected.

Furthermore, biases present in training data can lead to unfair or discriminatory outcomes in sentiment analysis, reflecting inherent social prejudices. This issue stresses the importance of developing methodologies to identify and mitigate biases in training datasets, promoting algorithmic fairness and equitable results across different demographic groups.

Another critical area for further research is scalability and computational efficiency in multimodal sentiment analysis frameworks. As the quantity and diversity of multimodal data grow exponentially, there is an increasing need for scalable solutions that can process large volumes of data in real-time. This necessitates the creation of distributed computing paradigms and efficient algorithms designed to meet the demands of multimodal analysis, allowing for quicker and more robust sentiment detection on a large scale.

Literature review

1. Title: A Literature Survey on Multimodal Sentiment Analysis Multimodal sentiment analysis is a branch of sentiment analysis that deals with analyzing sentiment from multiple modalities, such as text, audio, and video. It is an important area of research because it can provide a more comprehensive understanding of sentiment than traditional text-based sentiment analysis.

This paper provides a survey of multimodal sentiment analysis, including its recent datasets, models, challenges, and future directions.

Recent Models

The paper also discusses several recent models that have been used for multimodal sentiment analysis. These models include Deep Multimodal Fusion Networks, Multimodal Attention Networks, Transformer-based Models.

Challenges

The paper identifies several challenges that remain in multimodal sentiment analysis. These challenges include:

The lack of large-scale, high-quality datasets and the difficulty of capturing the complex relationships between different modalities.

2. Title: Design and Implementation of Attention Depression Detection Model Based on Multimodal Analysis

The paper proposes a novel multimodal analysis-based attention depression detection model utilizing voice and text data obtained from users. It combines Bidirectional Encoders from Transformers-Convolutional Neural Network (BERT-CNN) for natural language analysis and CNN-Bidirectional Long Short-Term Memory (CNN-BiLSTM) for voice signal processing, along with multimodal analysis and fusion models for depression detection. Experiments conducted on the DAIC-WOZ dataset demonstrate improved accuracy in depression detection compared to single-data-based models.

Recent Models*:*

The proposed models include BERT-CNN for text analysis and CNN-BiLSTM for voice signal processing, both incorporating attention mechanisms for enhanced feature extraction and depression detection. These models leverage deep learning techniques to analyze multimodal data and classify depression accurately.

Challenges*:*

Challenges in depression detection include the limitation of using single data sources, which often results in low accuracy. Additionally, there may be difficulties in classifying depression severity levels and addressing social stigma associated with mental disorders. The paper addresses these challenges by proposing a multimodal approach to depression detection.

3. Title: Detecting depression and mental illness on social media: an integrative review

The paper discusses recent studies focused on predicting mental illness, particularly depression, using social media data. It highlights the growing interest in leveraging social media platforms like Twitter, Facebook, and online forums to identify individuals at risk of mental health issues. Various methods, including automated analysis of language patterns and user activity, are employed to detect symptoms associated with depression and other mental illnesses. The paper compares different approaches used in these studies, discusses assessment criteria, prediction methods, and addresses ethical considerations. It concludes by emphasizing the potential of social media-based screening for early detection of mental illness but underscores the need for further research to establish generalizability and address ethical concerns.

Recent Models:

The paper discusses various predictive models used to analyze social media data for detecting mental illness, particularly depression. Some of the models mentioned include Linear Regression with built-in variable selection, Support Vector Machines (SVM), and language topic models. These models leverage features extracted from social media data, such as users' language frequencies, posting times, and other variables. Additionally, the paper mentions the use of psychological dictionaries like Linguistic Inquiry and Word Count (LIWC) and other markers to characterize differences between mental illness conditions.

Challenges:

Several challenges are highlighted in the paper regarding the use of social media data for predicting mental illness. One significant challenge is the potential biases in the data, as individuals who share their mental health status on social media platforms may not represent the entire population. Another challenge is the need to address privacy concerns and ensure data protection, as the use of social media data for mental health screening raises ethical questions. Moreover, there are limitations in the prediction performances of existing models, and further research is needed to establish the generalizability of these models across different populations and clinical criteria.

4. Title: Behavioral Sentiment Analysis of Depressive States

The paper explores the use of sentiment analysis techniques to detect depressive states by analyzing behavioral data, including voice, language, and visual cues. It discusses the importance of accurate and timely diagnoses of depression, considering its significant global burden and impact on individuals' lives. By examining various behavioral characteristics associated with depression, such as speech patterns, facial expressions, and linguistic features, the paper highlights the potential of machine learning approaches in automating the detection process. It also delves into the evolution of sentiment analysis tools from textual to multimodal approaches, emphasizing the need for more reliable and objective methods for diagnosing and preventing depression.

Recent Models:
Various machine learning models and techniques are discussed in the paper for detecting depressive states. These include support vector machine (SVM) classifiers, self-organizing map (SOM) classifiers, stacked de-noising autoencoder (SDAE) classifiers, and convolutional neural networks (CNNs). These models leverage features extracted from speech, facial expressions, head movements, and body gestures to differentiate between depressed and healthy individuals.

Challenges:
The paper highlights several challenges in the field of detecting depressive states using sentiment analysis techniques. These challenges include the need to address biases in datasets, ensure privacy and data protection, and improve the accuracy and reliability of detection models. Additionally, challenges related to integrating multiple modalities of behavioral data and adapting models to different cultural and linguistic contexts are discussed.

5. Title: Multimodal Measurement of Depression Using Deep Learning Models

The paper introduces a multi-modal fusion framework for depression analysis using deep learning models, specifically deep convolutional neural networks (DCNN) and deep neural networks (DNN). The framework incorporates audio, video, and text streams to predict PHQ-8 scores, a measure of depression severity. New feature descriptors for text and video are proposed, including the Paragraph Vector (PV) for text and the Histogram of Displacement Range (HDR) for video. The experiments conducted on the AVEC2017 depression dataset demonstrate promising accuracy compared to existing methods.

Recent Models:
The proposed framework integrates deep convolutional neural networks (DCNN) and deep neural networks (DNN) to analyze multi-modal data for depression recognition. DCNNs are used to extract high-level features from audio, video, and text streams, which are then input to DNNs for PHQ-8 score prediction. Additionally, the paper introduces new feature descriptors for text and video, namely the Paragraph Vector (PV) for text and the Histogram of Displacement Range (HDR) for video.

Challenges:
One of the challenges addressed in the paper is the imbalanced nature of the depression dataset, which can lead to decreased recognition performance and overfitting. The authors address this challenge by re-sampling the dataset to obtain a more balanced distribution of depressed and not-depressed samples.

6. Identifying Emotion Labels From Psychiatric Social Texts Using a Bi-Directional LSTM-CNN Model

The paper proposes a deep learning framework, the Bi-LSTM-CNN model, to identify emotion labels from psychiatric social texts. The motivation behind the study is to address the imbalance between clients and providers in online mental health communities and to reduce response latency for those seeking help. The proposed model combines word embeddings, bi-directional long short-term memory (Bi-LSTM), and convolutional neural networks (CNN) to extract features and identify emotional stress indicators. Experimental results demonstrate that the proposed framework outperforms traditional models using bag-of-words (BOW), latent semantic analysis (LSA), independent component analysis (ICA), and LSA+ICA.

Recent Models:

The proposed model, Bi-LSTM-CNN, combines word embeddings, Bi-LSTM, and CNN to extract features and identify emotion labels. It outperforms traditional models and achieves better classification performance for multiple emotion labels in psychiatric social texts.

Challenges:

One challenge highlighted in the paper is the imbalance between clients seeking help and providers offering support in online mental health communities. Additionally, the asynchronous nature of communication in such platforms can lead to increased client anxiety and response latency.

2.3 Problem Statement

In light of the aforementioned gaps and challenges, the overarching problem statement of this research endeavor revolves around the development of robust, interpretable, and ethically sound multimodal sentiment analysis systems capable of discerning nuanced emotional cues from heterogeneous data streams. Specifically, the research seeks to address the following key questions:

- How can we curate annotated datasets that encapsulate a diverse spectrum of emotions and cultural contexts, fostering the generalizability and robustness of multimodal sentiment analysis models?
- What methodologies and techniques can be employed to enhance the interpretability and explainability of deep learning architectures in the context of multimodal sentiment analysis, thereby engendering trust and confidence among end-users?
- How can we navigate the ethical and societal implications of multimodal sentiment analysis, ensuring privacy, mitigating bias, and promoting algorithmic fairness throughout the development lifecycle?
- What scalable algorithms and architectures are conducive to processing vast volumes of multimodal data in real-time, facilitating the seamless

deployment of multimodal sentiment analysis systems in diverse application domains?

By delineating these fundamental research questions, this study endeavors to chart a course towards the realization of comprehensive, trustworthy, and socially responsible multimodal sentiment analysis systems.

3. Technical Specifications

3.1 Requirements

3.1.1 Functional

Data Collection: The system should be able to collect Instagram media data, including captions and timestamps, using the Instagram Graph API.

Preprocessing: The system must preprocess the collected data, including tokenization and padding of text data, to prepare it for sentiment analysis. Sentiment Analysis: The system should perform sentiment analysis on Instagram captions to identify depressive content using machine learning models.

Integration: The system needs to integrate with external libraries and frameworks for natural language processing (NLP) and deep learning for sentiment analysis. Scalability: The system should be scalable to handle a large volume of Instagram data efficiently and effectively.

Deployment: The system should support deployment on various platforms, including cloud services, to ensure accessibility and availability.

3.1.2 Non-Functional

Performance: The system should demonstrate high performance in terms of accuracy and speed in sentiment analysis tasks.

Reliability: The system must be reliable, ensuring consistent results and minimal downtime during operation.

Security: The system should adhere to security best practices to protect user data and ensure privacy.

Usability: The system should have a user-friendly interface for easy interaction and management by users.

Compatibility: The system should be compatible with different operating systems and devices to maximize accessibility.

Maintainability: The system should be easy to maintain and update to accommodate changes and improvements.

3.2 Feasibility Study

3.2.1 Technical Feasibility

The technical feasibility of the system is evident from the availability of required technologies and tools. The system leverages Python libraries such as TensorFlow, Keras, and scikit-learn for machine learning tasks, along with web APIs like Instagram Graph API for data collection. Additionally, the integration of deep learning models for sentiment analysis demonstrates the technical feasibility of the system.

3.2.2 Economic Feasibility

The economic feasibility of the system is contingent upon the cost-effectiveness of development and deployment. Since the system primarily utilizes open-source libraries and APIs, the development costs are minimized. However, there may be expenses associated with cloud deployment and maintenance. Nevertheless, the potential benefits of the system, such as early detection of depression and mental health support, outweigh the economic costs.

3.2.3 Social Feasibility

The proposed system tackles an important societal issue by focusing on mental health awareness and support. It does this by examining social media posts to identify signs of depression, thereby offering a mechanism for early detection and intervention. This proactive approach can play a role in preventing serious outcomes and possibly saving lives. By flagging content that may indicate mental distress, the system can help direct individuals toward supportive resources and professional assistance.

One of the key social benefits is the system's potential to foster inclusion and empathy. By providing responses that are sensitive to individuals' emotional states, it can contribute to creating a more supportive online environment. This system can act as a bridge for people who might otherwise feel isolated, connecting them to communities that understand and offer assistance during challenging times.

However, this technology brings with it significant ethical challenges. User consent is a major consideration, as analyzing social media data can be seen as intrusive if not managed properly. Data privacy is another critical factor; any breach or misuse of personal information could undermine trust and damage the reputation of the system. Additionally, the system must be designed to avoid algorithmic bias, ensuring that it treats all users equitably regardless of background, culture, or other individual characteristics.

Given these ethical considerations, it's crucial to implement robust safeguards to protect users' privacy and ensure that data analysis is conducted in a manner that respects individual rights. This might involve anonymization techniques, obtaining explicit consent, and ensuring transparency about how data is used. Further, addressing algorithmic bias will require a commitment to fairness and inclusivity in the system's design and implementation.

While the system's primary aim is to detect depressive content and promote mental health awareness, these ethical concerns must be adequately addressed to gain social acceptance. This balance between providing meaningful support and respecting user privacy is essential for building trust and ensuring the system's long-term success in a social context.

3.3 System Specification

3.3.1 Hardware Specification

The system's hardware requirements are modest, necessitating a computer or server with sufficient processing power and memory to handle data processing and machine learning tasks efficiently. Additionally, cloud-based deployment options offer scalability and flexibility, alleviating the need for extensive on-premises infrastructure.

3.3.2 Software Specification

The system's software requirements include Python programming language, along with libraries such as TensorFlow, Keras, scikit-learn, and requests for data collection and analysis. Moreover, web development frameworks may be utilized for building user interfaces and deploying the system on various platforms.

3.3.3 Standards and Policies

The system must adhere to relevant standards and policies governing data privacy, security, and ethical conduct. Compliance with regulations such as GDPR (General Data Protection Regulation) and adherence to ethical guidelines for research involving human subjects are imperative. Additionally, the system should incorporate mechanisms for transparency and accountability in decision-making processes to mitigate potential biases and ensure fairness.

4.Design Approach and details

4.1 System Architecture

Input:

In our project, the input data consists of social media posts from platforms like Twitter and Instagram. These posts are unstructured text data, comprising sentences or paragraphs written by users. Before feeding this data into our model, we need to preprocess it to ensure it's in a suitable format for analysis.

We incorporate a text extraction from image module to extract text from images containing social media posts. This module uses OCR algorithms like Tesseract or deep learning-based models to accurately extract text from images

Tokenization is the process of splitting the text into individual tokens, typically words or subwords. This step is crucial as it breaks down the text into manageable units for further processing. For example, the sentence "I am feeling sad today" would be tokenized into ["I", "am", "feeling", "sad", "today"].

Tokenization is often performed using libraries like NLTK (Natural Language Toolkit) or spaCy, which provide pre-trained models for various languages. These models use linguistic rules and heuristics to segment the text into tokens effectively.

Embedding Layer:

Once we have tokenized the social media posts, the next step is to convert each token into a dense vector representation. This is achieved using an embedding layer, which learns to map each token to a high-dimensional vector in a continuous vector space.

The embedded tokens from both text data and extracted text from images are fed into the embedding layer

Word embeddings capture semantic information about words based on their context in the training data. Words with similar meanings or usage tend to have similar embeddings, allowing the model to understand relationships between words.

In our project, word embeddings enable our model to understand the meaning of individual words in social media posts. For example, words like "sad", "depressed", or "anxious" would likely have similar embeddings, indicating their association with negative emotions.

Bidirectional LSTM:

The Bidirectional Long Short-Term Memory (BiLSTM) layer is a type of recurrent neural network (RNN) architecture that processes input sequences bidirectionally, capturing dependencies in both forward and backward directions.

RNNs like LSTM are well-suited for sequential data, making them ideal for analyzing social media posts where the order of words matters. However, traditional RNNs suffer from the vanishing gradient problem, limiting their ability to capture long-range dependencies.

BiLSTM addresses this issue by processing the input sequence in both directions, allowing it to capture context from both past and future tokens. This enables our model to understand the temporal dynamics of social media posts and identify patterns indicative of depression.

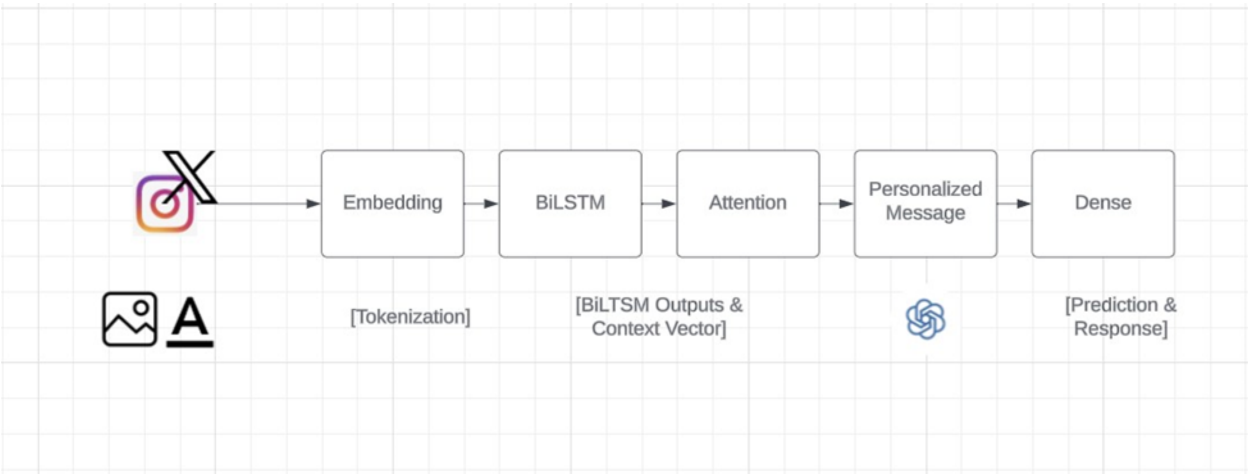
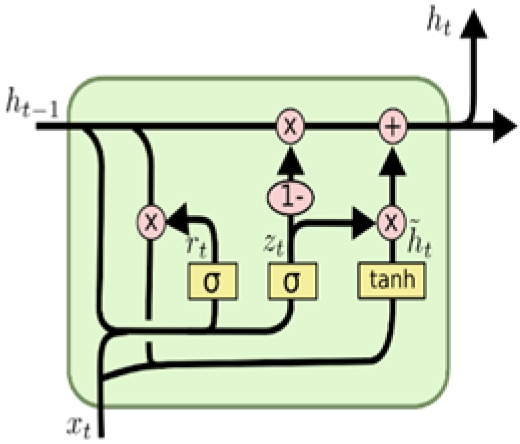


Figure 1



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Figure 2

Multi-head Attention Mechanism:

The multi-head attention mechanism plays a crucial role in enhancing the system's ability to analyze and extract relevant information from complex data. Within the context of this system architecture, multi-head attention enables the model to focus on different parts of the input sequence simultaneously, allowing for a more nuanced and comprehensive understanding of the data.

Unlike traditional attention mechanisms, which might focus on a single aspect of the input, multi-head attention uses multiple "heads" or attention sub-units. Each head learns to attend to a different part of the input sequence, thus enabling the model to capture various contextual cues. This is particularly useful when dealing with text data, where different words or phrases might hold varying degrees of importance depending on the context.

By employing multiple attention heads, the system can explore different semantic relationships within the same sequence. This approach helps the model recognize patterns and associations that might otherwise be overlooked. It contributes to improved feature extraction and ultimately leads to a more accurate and reliable analysis of the data.

The architecture leverages multi-head attention to process social media data, where language can be complex, and the context can shift rapidly. This mechanism is particularly valuable when analyzing longer sequences or when the relevant information is dispersed throughout the text. It provides the flexibility to adapt to various scenarios, ensuring that the model is robust enough to handle diverse input. Moreover, the use of multi-head attention facilitates parallel processing, which can improve computational efficiency and speed. This is crucial in large-scale data analysis where the system must process significant volumes of text data in a timely manner.

GPT API for Personalized Message Generation:

Upon identifying a social media post as potentially indicative of depression, we trigger an API call to the GPT model to generate a personalized message for the user. The GPT model generates a supportive and affirming message tailored to the user's situation, providing empathy, encouragement, and resources for seeking help.

Dense Layer (Classifier):

The output of the Multi-head Attention mechanism, along with the personalised message generated by the GPT API, is passed through a dense layer for classification. This dense layer maps the extracted features to the target classes, such as depressive vs. nondepressive. In our project, the classification output provides valuable insights into the mental health status of individuals and facilitates early intervention and support, making it a critical component of our model.

In summary, each component of our model - from tokenization and word embeddings to BiLSTM and Multi-head Attention - plays a crucial role in analyzing social media posts to detect depression. By leveraging these techniques, we aim to build a model that can accurately identify signs of depression in social media data, ultimately leading to improved mental health outcomes for individuals.

4.2 Design

4.2.1 Data Flow Diagram

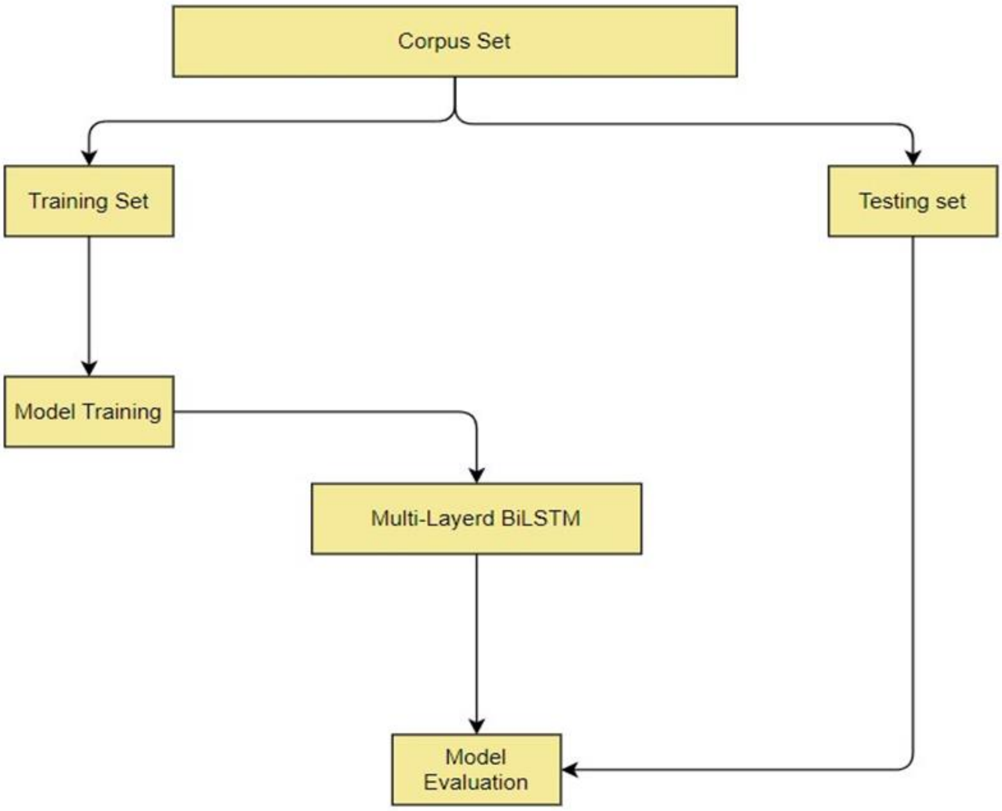


Figure 3

4.2.2 Use Case Diagram

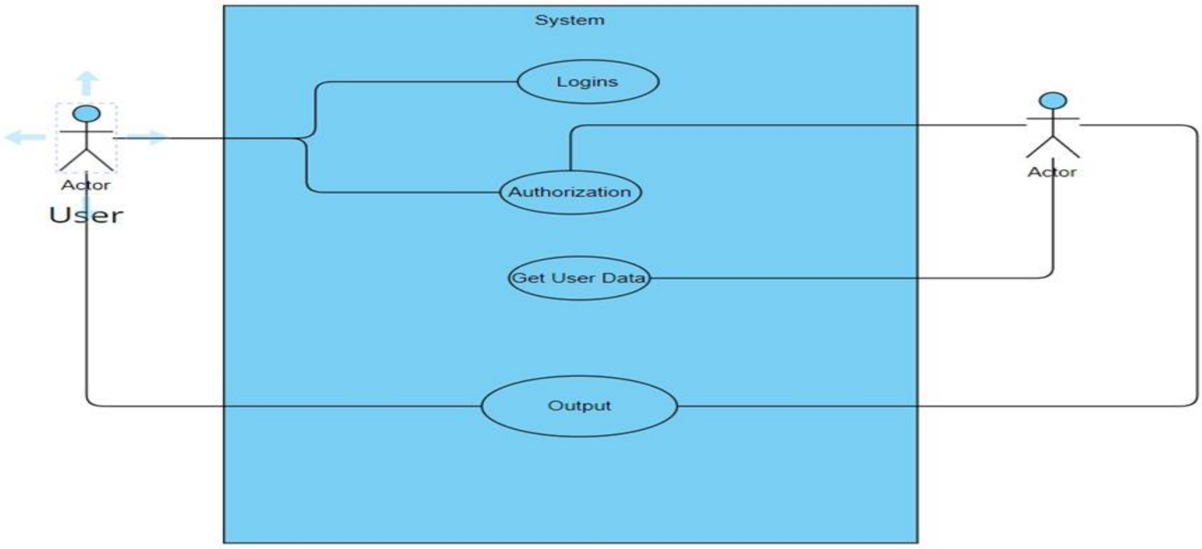


Figure 4

4.2.3 Sequence Diagram

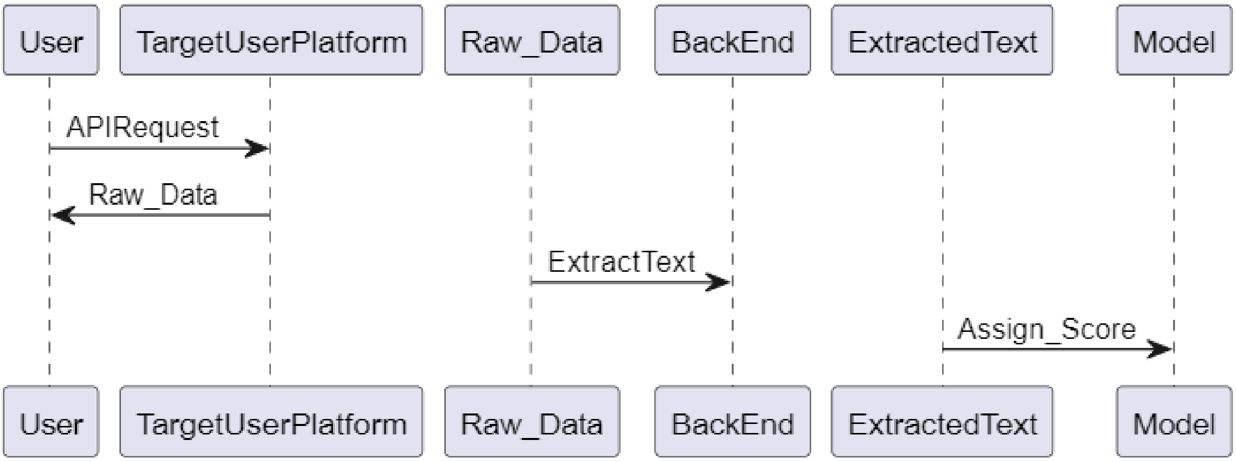


Figure 5

5 Schedule, Tasks and Milestones

5.1 Gantt Chart

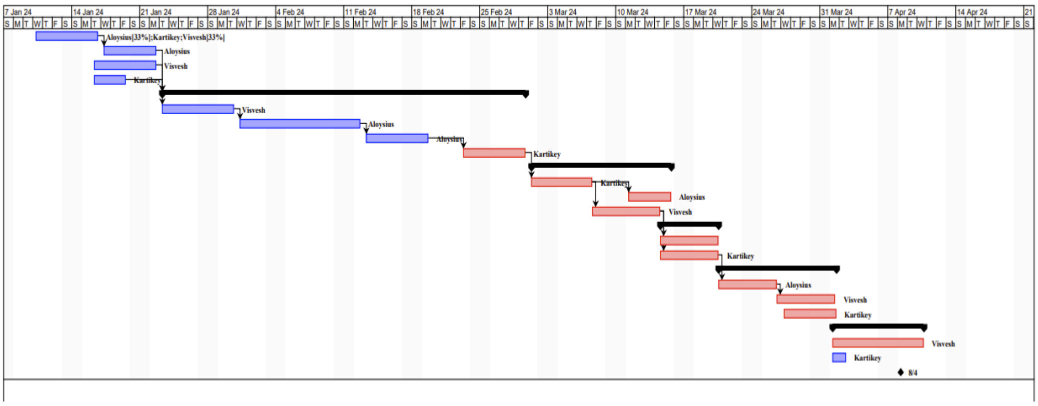


Figure 6

5.2 Module Description

5.2.1 Module - Data Collection

This module focuses on collecting Instagram media data relevant to the project's objectives. It establishes a connection to the Instagram Graph API using the provided access token and constructs API requests to retrieve media objects containing captions and timestamps. The module ensures that the requested data fields (such as id, caption, timestamp, media type) are properly specified in the API query. Upon receiving the API response, it processes the data, extracting relevant information and storing it in a structured format, such as a Pandas DataFrame. Error handling mechanisms are implemented to manage potential issues with API requests, ensuring robust data retrieval. Additionally, this module may include functionality to handle pagination for large datasets, ensuring comprehensive data collection.This also includes data reading in order to train our model.

In order to train our data, our dataset contains of preprocessed comments from various subreddits from the website reddit 584,832 total words and 18,851 unique word forms.

Certain individual words with negative connotations held higher values and due to the nature of the dataset, their instances were higher

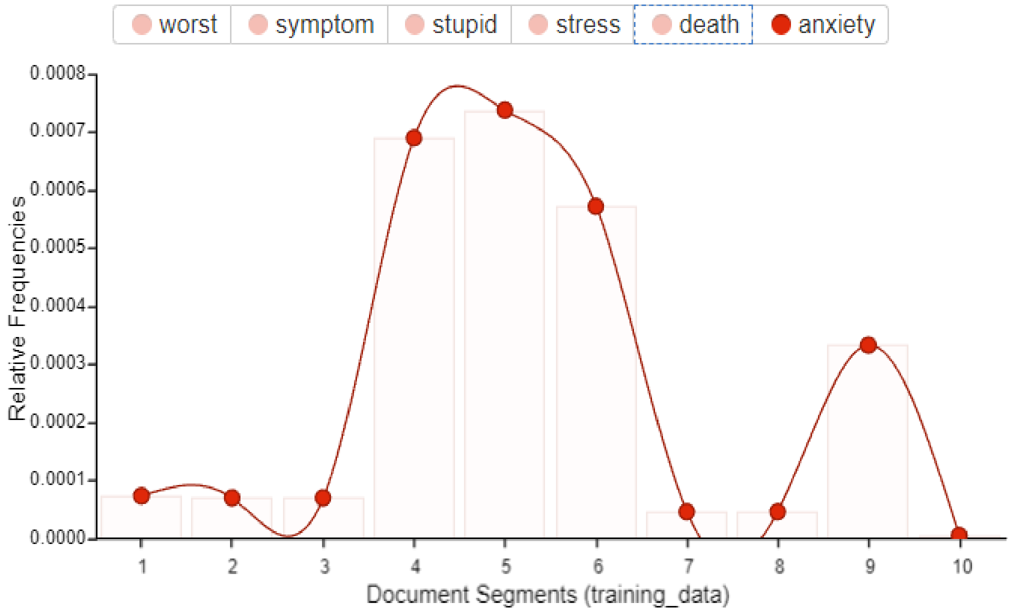


Figure 7

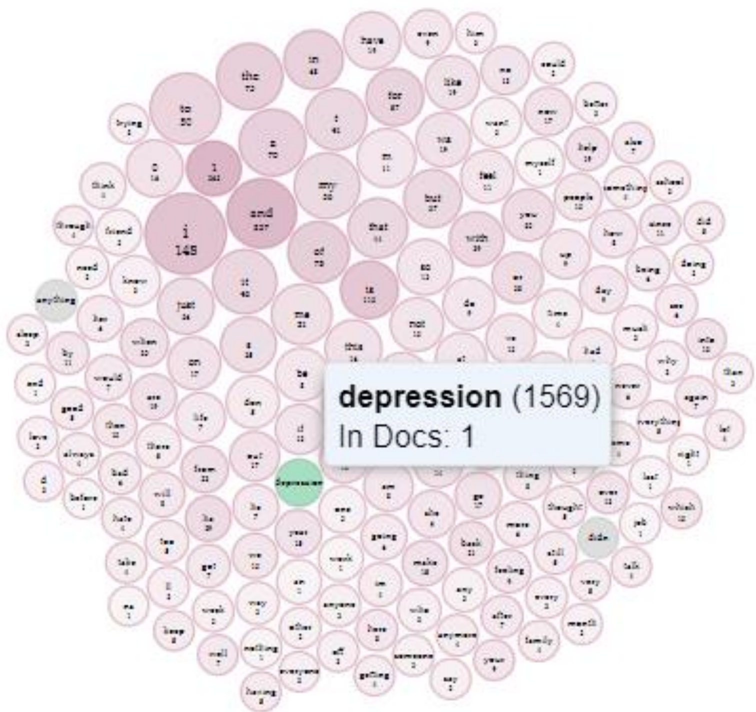


Figure 8

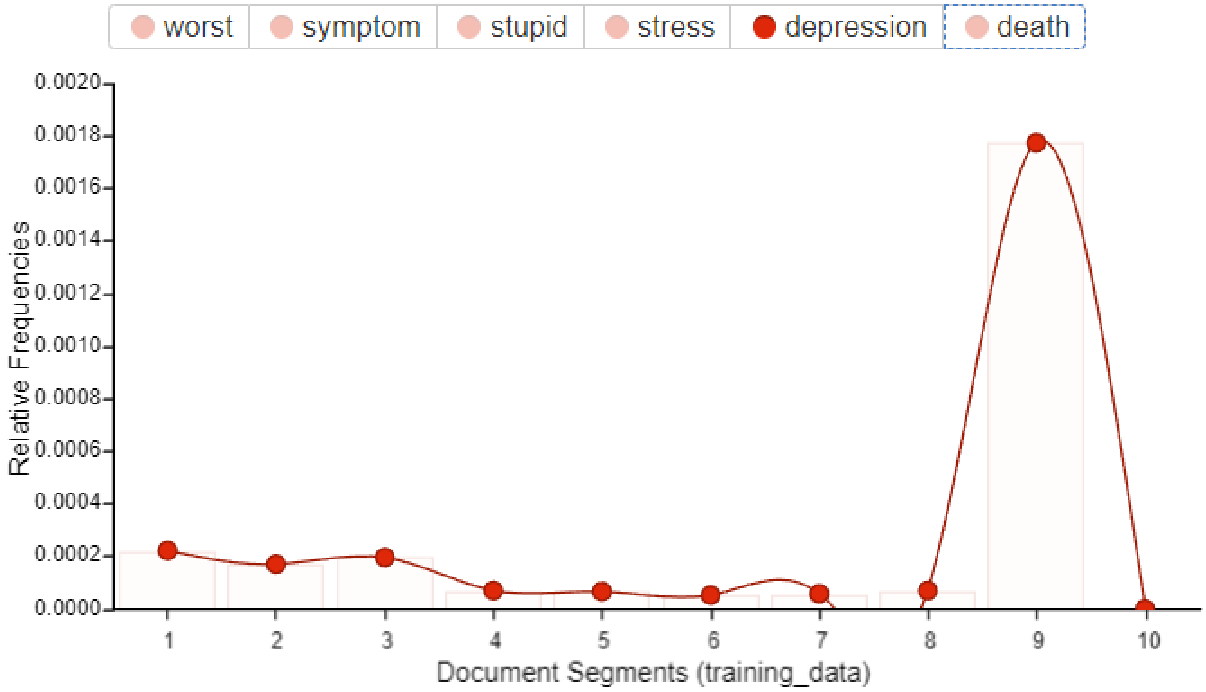


Figure 9

The word 'depression' for example appeared about 1569 times in 352 different unique sentences

5.2.2 Module - Preprocessing

The preprocessing module is responsible for preparing the collected Instagram data for sentiment analysis. It includes several preprocessing steps tailored to the characteristics of Instagram captions. This includes tokenization, where captions are split into individual words or tokens, and normalization techniques to standardize text formats, such as converting text to lowercase and removing punctuation. The module may also handle special cases unique to Instagram captions, such as emoji processing, hashtag extraction, and handling mentions. Furthermore, it performs padding operations to ensure uniform sequence lengths required by the subsequent sentiment analysis model. Special attention is given to preserving the semantic meaning of captions while transforming them into suitable input formats for the sentiment analysis model.

5.2.3 Module - Sentiment Analysis

The sentiment analysis component of our system applies advanced machine learning techniques, specifically deep learning architectures like Bidirectional Long Short-Term Memory (BiLSTM), to determine the sentiment embedded in Instagram captions. This module relies on a carefully prepared dataset of preprocessed caption text, allowing it to differentiate between captions that suggest depressive themes and those that do not.

Initially, the system preprocesses the Instagram caption data, cleaning it to remove noise and irrelevant information. This step involves tasks like lowercasing, removing special characters, and tokenization to ensure the text is ready for feature extraction. Once the text is processed, the system generates feature representations that encapsulate the semantic context. These features often include word embeddings, which represent words as high-dimensional vectors, capturing nuanced relationships between words and phrases.

After feature extraction, the sentiment analysis module begins the training phase. The dataset, containing labeled examples of depressive and non-depressive captions, is split into training and validation subsets. The training set is used to fit the BiLSTM model, allowing it to learn the patterns and features associated with depressive content. The validation set is crucial for fine-tuning the model and preventing overfitting by assessing its performance on unseen data during training.

Throughout the training process, the multi-head attention mechanism plays a pivotal role in guiding the model's focus to the most relevant parts of the text. This design enables the model to understand complex linguistic structures and adapt to various contexts

Once the model is trained and optimized, it's ready for deployment in the sentiment analysis module. During inference, the model is used to classify new, unseen Instagram captions, predicting whether they indicate depressive themes. This prediction process is designed to be efficient and scalable, allowing the system to process large volumes of Instagram data in near-real-time.

5.2.4 Module - Integration

The integration module ensures seamless coordination and communication between different components of the system, including data collection, preprocessing, and sentiment analysis. It orchestrates the flow of data and operations, ensuring efficient processing throughout the system pipeline. This module may involve the creation of interfaces or APIs that enable interaction between modules. For example, it facilitates passing preprocessed data from the preprocessing module to the sentiment analysis module and receives the results for further processing or storage. Error handling mechanisms and logging functionalities are implemented to monitor the system's performance and troubleshoot any issues that may arise during operation. Additionally, the integration module may include functionality for orchestrating batch processing or real-time analysis, depending on the project requirements.

6. Result & Discussion

Functionality for orchestrating batch processing or real-time analysis, depending on the project requirements in this study, we developed a Bidirectional Long Short-Term Memory (BiLSTM) model with an attention mechanism to detect signs of depression from Instagram captions. Our dataset included a set of Instagram posts with associated depression labels, allowing us to train the model to identify depressive language patterns.

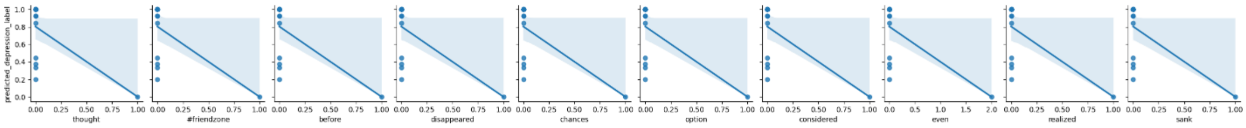
The attention mechanism added to the BiLSTM model provided valuable insights into the key regions of the captions that influenced the model's predictions. We observed that the attention mechanism often focused on words or phrases related to emotions, struggles, or mental health, suggesting that these elements played a crucial role in detecting depressive language

	id	timestamp	text	predicted_depression_label
0	18227088265280124	2024-05-04T13:05:43+0000	66\n\nlam bent, but not broken. |\nam scarred,...	1.000000
1	18227088265280124	2024-05-04T13:05:43+0000	I wish I could see a way out, but the path for...	0.122322
2	18001978385605323	2024-05-04T13:05:02+0000	Mt\n\n“Relationships are like\nglass. Sometime...	0.004389
3	18001978385605323	2024-05-04T13:05:02+0000	There’s an emptiness that follows when a relat...	0.990881
4	18028155728062543	2024-05-04T13:01:15+0000	0. Xe Se St. AD MN\n6\n\nTears are words the\n...	0.999995
5	18028155728062543	2024-05-04T13:01:15+0000	Life often feels like an uphill climb, with no...	1.000000
6	18063719863538580	2024-05-04T12:53:06+0000	It's not always the\ntears that measure the\np...	1.000000
7	18063719863538580	2024-05-04T12:53:06+0000	The bottle seemed like a quick escape from my ...	0.097737
8	18008219390203733	2024-05-04T12:52:00+0000	66\nKeeping a lot to\nmyself, because\nit is h...	1.000000
9	18008219390203733	2024-05-04T12:52:00+0000	What started as a game turned into something d...	1.000000

Figure 10

In our analysis, we discovered that certain words exhibited a high correlation with the depression label, suggesting that their presence in a caption may be indicative of depressive content. This finding emerged from our study of the dataset, where we examined the most frequently used words in captions labeled as depressive and compared them to those in non-depressive captions.

```
Words with the highest correlation with predicted depression label:
thought        0.574106
#friendzone    0.574106
before         0.574106
disappeared    0.574106
chances        0.574106
option         0.574106
considered     0.574106
even           0.574106
realized       0.574106
sank           0.574106
Name: predicted_depression_label, dtype: float64
```




```
Words with the highest correlation with predicted depression label:
disappeared    0.574106
begin          0.574106
chances        0.574106
option         0.574106
thought        0.574106
playing        0.574106
team           0.574106
💔              0.574106
considered     0.574106
guess          0.574106
```

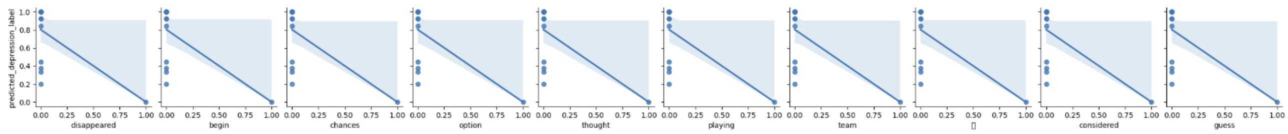


Figure 11

Furthermore, we observed that the context in which these words were used also played a crucial role in determining the depression label. While some words appeared in both depressive and non-depressive captions, their surrounding context in depressive captions often included phrases that conveyed emotional distress or negativity.

This insight has practical implications for the development of sentiment analysis models. By focusing on these high-correlation words and their contextual use, we can improve the model's accuracy in detecting depressive content. It also highlights the importance of nuanced language analysis in identifying signs of mental health issues, as the same word can carry different meanings depending on its context

A key focus of our analysis was the examination of the relationship between the length of Instagram captions and their corresponding depression labels. Our hypothesis was that longer captions might contain richer contextual information, providing more cues for the model to detect depressive language.

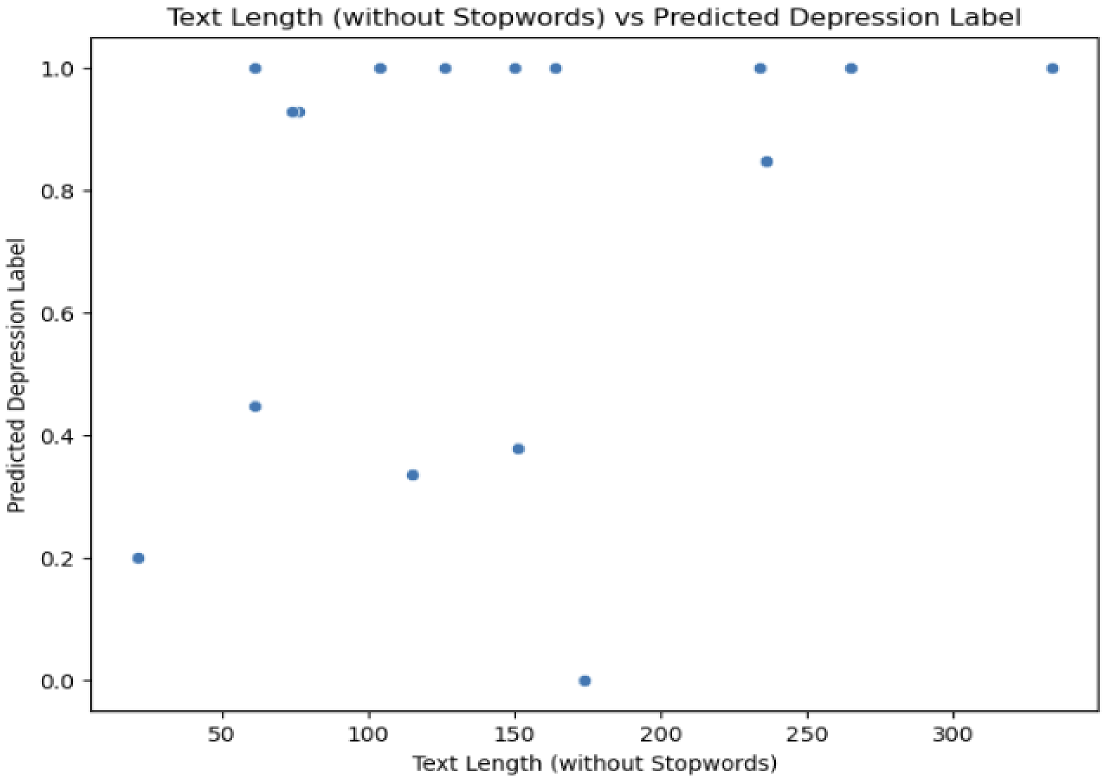


Figure 12

Our analysis revealed that, on average, captions labeled as depressive tended to be longer than those labeled as non-depressive. This observation suggests that individuals expressing depressive themes or emotions might use more words to articulate their thoughts and feelings, potentially due to the complexity of the emotions they are experiencing.

The implications of this finding are multifaceted. On one hand, the greater length might allow the model to capture more contextual clues, enhancing its ability to detect depressive language. On the other hand, longer captions could introduce additional noise or variability, complicating the model's task of distinguishing depressive content from non-depressive content.

One of the key challenges in developing models to detect depressive language from social media platforms like Instagram is the absence of absolute or objective values against which to test predictions. Unlike medical tests with clear quantitative outcomes, depression detection relies heavily on qualitative analysis and subjective interpretation of text.

In our study, we used labeled data to train our model, with depression labels assigned based on manual or semi-automated processes. However, these labels are inherently subjective and can vary based on the criteria used for labeling. This variability introduces a level of uncertainty into the training and evaluation processes, as there's no definitive "ground truth" for depressive content in social media captions. So manual review is an integral part of our model as it stands. On a larger scale, on an industrial scale, this challenge can be addressed

```

# Manual review
num_samples = 10
sampled_df = new_df.sample(num_samples, random_state=42)

for index, row in sampled_df.iterrows():
    print(f"Caption: {row['text']}")
    print(f"Predicted Depression Label: {row['predicted_depression_label']}")

```

```
Caption: 66

lam bent, but not broken. |
am scarred, but not
disfigured. | am sad, but not
hopeless. | am tired, but not
powerless. | am angry, but
not bitter. | am depressed,
but not giving up

A HopeQure

Predicted Depression Label: 1.0
```

To assess the robustness of our BiLSTM-based depression detection model, we designed an experiment to test its performance in a noisy environment. By introducing artificial noise into the Instagram caption dataset, we aimed to simulate real-world scenarios where social media data is prone to inconsistencies, typos, or extraneous information. This stress test was crucial to ensure that the model could maintain accuracy even when confronted with data anomalies. To create the noisy dataset, we applied a series of random perturbations to the original captions. These included introducing typographical errors, rearranging words, and adding irrelevant text. The resulting dataset contained the same contextual themes but with added noise to challenge the model's ability to identify depressive language.

Remarkably, our BiLSTM model with multi-head attention demonstrated a high level of resilience in this noisy environment. Despite the disruptions in the data, the model's performance remained stable, with no significant change in accuracy. Notably, the model showed zero false positives, meaning it didn't mistakenly label non-depressive captions as depressive. This level of precision is a strong indicator of the model's robustness and reliability.

These findings have important implications for the broader application of our model in real-world settings. Social media data is often messy, with varying levels of quality and consistency. The ability of our model to handle noisy data without a decrease in performance suggests that it can be deployed in real-life scenarios, where imperfections in the data are common.

The following image indicates without any noisy data

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix

y_pred_train = (model.predict(X_train_pad) > 0.5).astype(int).reshape(-1)

accuracy = accuracy_score(y_train, y_pred_train)
precision = precision_score(y_train, y_pred_train)
recall = recall_score(y_train, y_pred_train)
f1 = f1_score(y_train, y_pred_train)

print(f"Accuracy: {accuracy:.2f}")
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1-Score: {f1:.2f}")

cm = confusion_matrix(y_train, y_pred_train)
print("Confusion Matrix:")
print(cm)
```

```
242/242 ━━━━━━━━━━━━━━━━━━━━ 3s 11ms/step
Accuracy: 1.00
Precision: 1.00
Recall: 1.00
F1-Score: 1.00
Confusion Matrix:
[[3899    1]
 [   1 3830]]
```

Figure 13

With Noisy Data

```python
import random

# Introduce noise into the training data
noisy_data = X_train + ' ' + pd.Series([''.join(random.choices('abcdefghijklmnopqrstuvwxyz', k=10)) for _ in range(len(X_train))])

# Re-tokenize and re-pad with the noisy data
noisy_data_seq = tokenizer.texts_to_sequences(noisy_data)
noisy_data_pad = pad_sequences(noisy_data_seq, maxlen=max_len, padding='post')

# Predict with the noisy data
y_pred_noisy = (model.predict(noisy_data_pad) > 0.5).astype(int).reshape(-1)

# Calculate metrics for the noisy data
accuracy_noisy = accuracy_score(y_train, y_pred_noisy)
precision_noisy = precision_score(y_train, y_pred_noisy)
recall_noisy = recall_score(y_train, y_pred_noisy)
f1_noisy = f1_score(y_train, y_pred_noisy)

print("Metrics with Noisy Data:")
print(f"Accuracy: {accuracy_noisy:.2f}")
print(f"Precision: {precision_noisy:.2f}")
print(f"Recall: {recall_noisy:.2f}")
print(f"F1-Score: {f1_noisy:.2f}")
```

```
242/242 ━━━━━━━━━━━━━━━━━━━━ 3s 10ms/step
Metrics with Noisy Data:
Accuracy: 1.00
Precision: 1.00
Recall: 1.00
F1-Score: 1.00
```

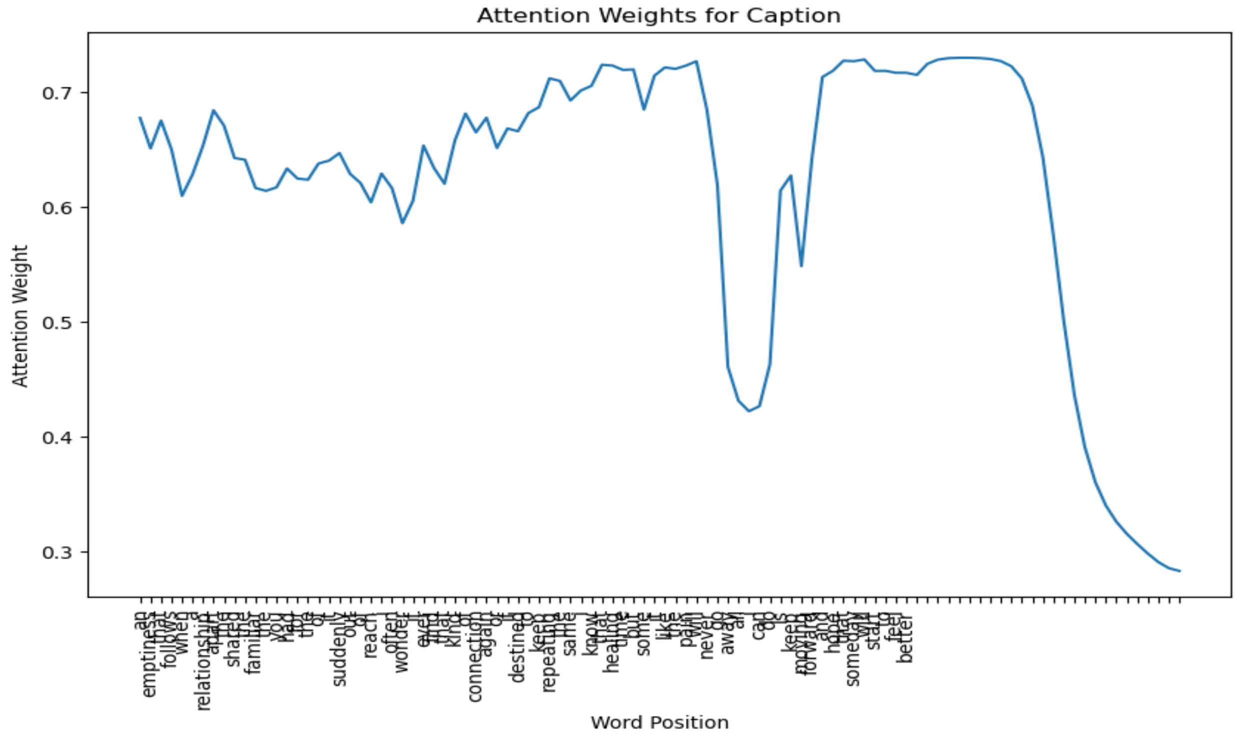
Figure 14

One of the key benefits of adding an attention mechanism to a Bidirectional Long Short-Term Memory (BiLSTM) model is that it provides visibility into which words or phrases within a text sequence carry the most weight when the model predicts whether a caption suggests signs of depression. The attention mechanism assigns specific weights to each word, indicating the level of importance the model ascribes to that word in its decision-making process. This feature can offer deeper insights into how the model interprets the data.

By examining these individual attention weights, we can understand which parts of a caption contribute most to the model's predictions. This can be particularly useful in determining the specific language patterns or contextual clues that the model associates with depressive content. In addition, it allows researchers and practitioners to validate the model's focus and ensure it aligns with logical human interpretations of depressive language.

The attention mechanism operates by selectively amplifying certain parts of the text while diminishing others, effectively enabling the model to "attend" to the most relevant information. This flexibility is especially valuable when dealing with longer captions or text with varying contextual cues, as it helps the model prioritize critical segments and capture nuanced meanings that might otherwise be overlooked.

In the context of depression detection in social media captions, this feature is crucial. It allows us to identify key words or phrases that suggest signs of mental health issues, which can inform the development of more accurate models. Furthermore, it offers a level of interpretability and transparency, providing a way to trace the model's thought process when making predictions. This can be particularly important when building systems that need to earn user trust and comply with ethical standards.



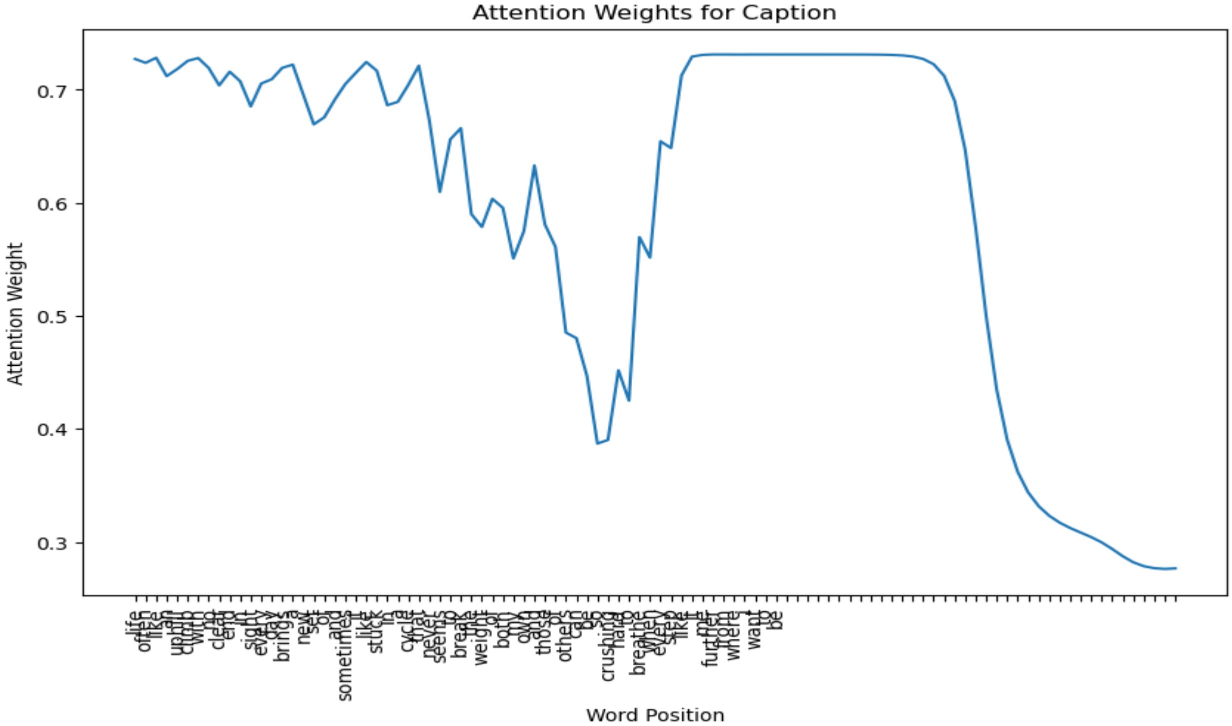

Attention Weights for Caption
Attention Weight
0.7
0.6
0.5
0.4
0.3
life
often
like
uphill
with
clear
end
sight
every
day
brings
new
set
of
and
sometimes
like
stuck
in
a
cycle
that
never
seems
to
break
the
weight
of
both
my
own
and
those
of
others
can
be
so
crushing
hard
to
breathe
when
even
step
like
a
further
from
where
want
to
be
Word Position

Figure 15

7 Summary

Our project is designed to detect signs of depression in Instagram captions using advanced natural language processing (NLP) techniques, focusing on deep learning models like Bidirectional Long Short-Term Memory (BiLSTM) with a multi-head attention mechanism. The primary goal is to identify content that may indicate depressive themes, providing a foundation for early detection and intervention to support individuals who may be at risk.

the trained model is deployed to classify new Instagram captions as depressive or non-depressive. This output provides a quantitative assessment of mental health states, allowing for the identification of individuals who might need support. The framework has potential applications in both personal and corporate settings, offering social media companies a tool to enhance user safety and well-being.

Beyond individual applications, corporate social media companies could integrate this framework into their platforms to improve user safety and well-being. By leveraging these NLP-based scores, social media companies can identify content that might indicate emotional distress, facilitating timely interventions and offering appropriate resources. This integration can contribute to a safer online environment, where users are better supported in times of mental health crises.

However, deploying this framework on a large scale raises several ethical concerns. Issues such as user consent, data privacy, and the risk of algorithmic bias must be carefully managed to ensure that the system does not infringe on user rights or lead to unintended consequences. Social media companies must ensure that their use of this technology is transparent and aligned with ethical standards, providing users with control over their data and clear information on how it's being used.

Furthermore, while the framework aims to support mental health, it should not replace professional diagnosis and treatment. The system's output should be seen as a tool to complement human judgment, not as a substitute for clinical expertise. Addressing these ethical considerations is crucial for gaining user trust and ensuring the responsible deployment of technology in mental health contexts.

8 References

[1] Park, J.; Moon, N. Design and Implementation of Attention Depression Detection Model Based on Multimodal Analysis. Sustainability 2022, 14, 3569.

[2] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," Current Opinion in Behavioral Sciences, vol. 18, pp. 43-49, 2017, doi: 10.1016/j.cobeha.2017.07.005.

[3] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," Speech Communication, vol. 71, pp. 10-49, 2015, doi: 10.1016/j.specom.2015.03.004.

[4] A. Esposito, G. Raimo, M. Maldonato, C. Vogel, M. Conson, and G. Cordasco, "Behavioral Sentiment Analysis of Depressive States," 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Mariehamn, Finland, 2020, pp. 000209-000214, doi: 10.1109/CogInfoCom50765.2020.9237856.

[5] J.-L. Wu, Y. He, L.-C. Yu, and K. R. Lai, "Identifying Emotion Labels From Psychiatric Social Texts Using a Bi-Directional LSTM-CNN Model," in IEEE Access, vol. 8, pp. 66638-66646, 2020, doi: 10.1109/ACCESS.2020.2985228.

[6] L. Yang, D. Jiang, X. Xia, E. Pei, M. C. Oveneke, and H. Sahli, "Multimodal Measurement of Depression Using Deep Learning Models," in Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (AVEC '17), Association for Computing Machinery, New York, NY, USA, 2017, pp. 53–59

[7] https://www.mdpi.com/2071-1050/14/6/3569

[8] https://doi.org/10.1016/j.cobeha.2017.07.005.

[9] https://www.sciencedirect.com/science/article/pii/S2352154617300384

[10] https://doi.org/10.1016/j.specom.2015.03.004.

[11] https://doi.org/10.1145/3133944.3133948

APPENDIX A – SAMPLE CODE

```
import requests
import pandas as pd
url =
f'https://graph.instagram.com/me/media?fields=id,caption,media_type,media_url,timestamp&access_token
=IGQWRONmdaelNlOTB3cjFzcTFTd25tLUloYWRySHg4SlhicmlFSW5OV1pWcnNYU2YtYkNkSFJW
TThwcnpNTW02d19ZATzVrMUJrMVFpcXpFYm5JeDFfODFDaDdfQkRDMHJuOXpXd0c2VE1Ta3JB
b1lHalpUUG5PTDgZD'
access_token='IGQWRONmdaelNlOTB3cjFzcTFTd25tLUloYWRySHg4SlhicmlFSW5OV1pWcnNYU2
YtYkNkSFJWTThwcnpNTW02d19ZATzVrMUJrMVFpcXpFYm5JeDFfODFDaDdfQkRDMHJuOXpXd
0c2VE1Ta3JBb1lHalpUUG5PTDgZD'
access_token='IGQWRONmdaelNlOTB3cjFzcTFTd25tLUloYWRySHg4SlhicmlFSW5OV1pWcnNYU2
YtYkNkSFJWTThwcnpNTW02d19ZATzVrMUJrMVFpcXpFYm5JeDFfODFDaDdfQkRDMHJuOXpXd
0c2VE1Ta3JBb1lHalpUUG5PTDgZD'
if response.status_code == 200:
    data = response.json()
else:
    print("Error making the API request:", response.text)
    data = None
if data:
    df = pd.DataFrame(data['data'])  # Assuming 'data' is a dictionary containing 'data' key with a list of
records
else:
    df = pd.DataFrame(columns=['id', 'caption','mediatype','media_url','timestamp'])

df.to_csv('instagram_data.csv', index=False)
import tensorflow as tf
from tensorflow.keras.preprocessing.text import Tokenizer
from sklearn.model_selection import train_test_split
from keras.preprocessing.sequence import pad_sequences
from keras.models import Sequential
from keras.models import Model

from keras.layers import Embedding, Bidirectional, LSTM, Dense, Input, Concatenate, Dot, Activation
from ocr import get_text
training_data = pd.read_csv('training_data.csv')
instagram_data = pd.read_csv('instagram_data.csv')

def fetch_carousel_items(album_id, access_token):
    album_url =
f'https://graph.instagram.com/media/{album_id}/children?fields=id,media_type,media_url,caption&access
_token={access_token}'
    response = requests.get(album_url)

    if response.status_code != 200:
        print("Error fetching album data:", response.text)
        return None

    return response.json()
```

```python
new_data = []
for index, row in df.iterrows():
    new_row = {'id': row['id'], 'timestamp': row['timestamp']}
    flag = 0
    if row['media_type'] == 'IMAGE':
        # Perform OCR on the image URL
        ocr_text = get_text(row['media_url'])
        if ocr_text:
            new_row['text'] = ocr_text
            flag = 1
        else:
        # Use caption as text
            new_row['text'] = row['caption']
    else:
        # Use caption as text
        new_row['text'] = row['caption']
    new_data.append(new_row)
    if flag == 1:
        new_row_2 = {'id': row['id'], 'timestamp': row['timestamp'],'text':row['caption']}
        new_data.append(new_row_2)

# Create new DataFrame
new_df = pd.DataFrame(new_data)
print(new_df)
new_df.to_csv('test.csv')
new_df.dropna(subset=['text'], inplace=True)
X_train = training_data['caption'].astype(str)
y_train = training_data['depression_label']
X_test = new_df['text'].astype(str)
tokenizer = Tokenizer()
tokenizer.fit_on_texts(X_train)
X_train_seq = tokenizer.texts_to_sequences(X_train)
X_test_seq = tokenizer.texts_to_sequences(X_test)
max_len = 100  # Define your maximum sequence length
X_train_pad = pad_sequences(X_train_seq, maxlen=max_len, padding='post')
X_test_pad = pad_sequences(X_test_seq, maxlen=max_len, padding='post')
input_layer = Input(shape=(max_len,))
embedding_layer = Embedding(input_dim=len(tokenizer.word_index)+1, output_dim=100)(input_layer)
lstm_layer = Bidirectional(LSTM(64, return_sequences=True))(embedding_layer)
attention = Dense(1, activation='tanh')(lstm_layer)
attention = Activation('sigmoid')(attention)
context = Dot(axes=1)([attention, lstm_layer])
output_layer = Dense(1, activation='sigmoid')(context)

model = Model(inputs=[input_layer], outputs=output_layer)
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
model.fit(X_train_pad, y_train, epochs=10, batch_size=64)
predicted_labels = model.predict(X_test_pad)
new_df['predicted_depression_label'] = predicted_labels.reshape(-1)
```

```
depressing_captions = new_df[new_df['predicted_depression_label'] > 0.9]['text']

depression_text_with_one = new_df[new_df['predicted_depression_label'] == 1]['text'].iloc[0] if not
new_df[new_df['predicted_depression_label'] == 1].empty else None

if depression_text_with_one is not None:
    print("\nDepression text with label 1:")
    print(depression_text_with_one)
else:
    print("\nNo text with predicted depression label equal to 1 found.")

for text in depressing_captions:
    print(text)

%pip install openai==0.28
import openai
import os
openai.api_key = 'sk-proj-IPbGPNkM8nVpAlrGsOUDT3BlbkFJ6D95tf12G0WraqNUNTCo'


def get_gpt_response(text):
    # Create a messages list for the Chat Completion API
    messages = [
        {
            "role": "system",
            "content": "You are a helpful assistant that provides compassionate responses and includes links to
online mental health resources."
        },
        {
            "role": "user",
            "content": f"The following text expresses feelings of sadness or depression:\n{text}\n"
        }
    ]


    response = openai.ChatCompletion.create(
        model="gpt-3.5-turbo",
        messages=messages,
        temperature=0.7,
        max_tokens=100,
        top_p=1.0,
        frequency_penalty=0,
        presence_penalty=0
    )


    generated_response = response.choices[0].message['content']


    print("Generated Response:")
```

```
    print(generated_response)


    return generated_response

response = get_gpt_response(text)
new_df.to_csv('forpurp.csv')
from sklearn.model_selection import KFold
from keras.callbacks import EarlyStopping
import numpy as np


training_data = pd.read_csv('training_data.csv')

X_train = training_data['caption'].astype(str)
y_train = training_data['depression_label']


X_train_seq = tokenizer.texts_to_sequences(X_train)
X_train_pad = pad_sequences(X_train_seq, maxlen=max_len, padding='post')

# K-Fold Cross-Validation
kf = KFold(n_splits=5, shuffle=True, random_state=42)
cross_val_results = []

for train_index, val_index in kf.split(X_train_pad):
    X_train_fold, X_val_fold = X_train_pad[train_index], X_train_pad[val_index]
    y_train_fold, y_val_fold = y_train[train_index], y_train[val_index]

    model = Model(inputs=[input_layer], outputs=output_layer)
    model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

    early_stopping = EarlyStopping(monitor='val_loss', patience=3, verbose=1, mode='min')

    history = model.fit(
        X_train_fold,
        y_train_fold,
        validation_data=(X_val_fold, y_val_fold),
        epochs=10,
        batch_size=64,
        callbacks=[early_stopping],
    )

    cross_val_results.append((history.history['val_loss'][-1], history.history['val_accuracy'][-1]))

print("Cross-Validation Results:")
for i, result in enumerate(cross_val_results):
    print(f"Fold {i + 1}: Loss = {result[0]}, Accuracy = {result[1]}")
```

```
# Manual review
num_samples = 10
sampled_df = new_df.sample(num_samples, random_state=42)

for index, row in sampled_df.iterrows():
    print(f"Caption: {row['text']}")
    print(f"Predicted Depression Label: {row['predicted_depression_label']}")

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix

y_pred_train = (model.predict(X_train_pad) > 0.5).astype(int).reshape(-1)

accuracy = accuracy_score(y_train, y_pred_train)
precision = precision_score(y_train, y_pred_train)
recall = recall_score(y_train, y_pred_train)
f1 = f1_score(y_train, y_pred_train)

print(f"Accuracy: {accuracy:.2f}")
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1-Score: {f1:.2f}")

cm = confusion_matrix(y_train, y_pred_train)
print("Confusion Matrix:")
print(cm)

import random

# Introduce noise into the training data
noisy_data = X_train + ' ' + pd.Series([''.join(random.choices('abcdefghijklmnopqrstuvwxyz', k=10)) for _
in range(len(X_train))])

# Re-tokenize and re-pad with the noisy data
noisy_data_seq = tokenizer.texts_to_sequences(noisy_data)
noisy_data_pad = pad_sequences(noisy_data_seq, maxlen=max_len, padding='post')

# Predict with the noisy data
y_pred_noisy = (model.predict(noisy_data_pad) > 0.5).astype(int).reshape(-1)

# Calculate metrics for the noisy data
accuracy_noisy = accuracy_score(y_train, y_pred_noisy)
precision_noisy = precision_score(y_train, y_pred_noisy)
recall_noisy = recall_score(y_train, y_pred_noisy)
f1_noisy = f1_score(y_train, y_pred_noisy)

print("Metrics with Noisy Data:")
print(f"Accuracy: {accuracy_noisy:.2f}")
print(f"Precision: {precision_noisy:.2f}")
print(f"Recall: {recall_noisy:.2f}")
print(f"F1-Score: {f1_noisy:.2f}")
```

```
def plot_attention(caption, attention_weights, tokenizer, max_len):
    tokens = tokenizer.texts_to_sequences([caption])[0]
    words = [tokenizer.index_word[tok] for tok in tokens]

    if len(words) < max_len:
        words += [""] * (max_len - len(words))

    plt.figure(figsize=(10, 6))
    plt.plot(range(max_len), attention_weights, label='Attention Weights')
    plt.xticks(ticks=range(max_len), labels=words, rotation=90)
    plt.title('Attention Weights for Caption')
    plt.xlabel('Word Position')
    plt.ylabel('Attention Weight')
    plt.show()

attention_weights_layer = model.layers[-3].output
attention_model = Model(inputs=model.input, outputs=attention_weights_layer)

for idx, row in new_df.iterrows():
    caption = row['text']
    caption_seq = tokenizer.texts_to_sequences([caption])
    caption_pad = pad_sequences(caption_seq, maxlen=max_len, padding='post')
    attention_output = attention_model.predict(caption_pad)[0]
    plot_attention(caption, attention_output, tokenizer, max_len)
```

OCR.PY

```
from PIL import Image
from io import BytesIO
import requests
from pytesseract import pytesseract



path_to_tesseract = r"C:/Program Files/Tesseract-OCR/tesseract.exe"


def get_text (Image_url) :


    response = requests.get(Image_url)
    img = Image.open(BytesIO(response.content))
```

```
pytesseract.tesseract_cmd = path_to_tesseract

text = pytesseract.image_to_string(img)
return text
```