Sentiment Analysis Using Text Classification on Social Media Post

by

Rida Mahmud Sadman
20301096
Asibur Rahman Marin
20301202
Sadman Sami
20301068
Snigdha Islam Nilima
21101213
Md Shadman Shakib
20301127

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

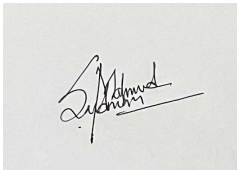
Department of Computer Science and Engineering
Brac University
June 2025

Declaration

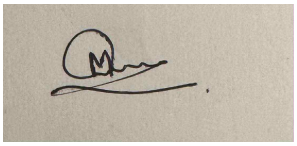
It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

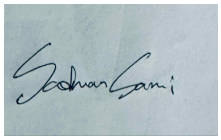
Student's Full Name & Signature:



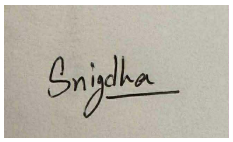
Rida Mahmud Sadman
20301096



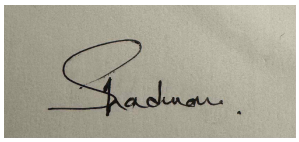
Asibur Rahman Marin
20301202



Sadman Sami
20301068



Snigdha Islam Nilima
211012134



Md Shadman Shakib
20301127

Approval

The thesis titled "Sentiment Analysis Using Text Classification on Social Media Post" submitted by

1. Rida Mahmud Sadman (20301096)
2. Asibur Rahman Marin (20301202)
3. Sadman Sami (20301068)
4. Snigdha Islam Nilima (21101213)
5. Md Shadman Shakib (20301127)

Of Summer, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in June, 2024.

Examining Committee:

Supervisor:
(Member)

Tanzim Reza
Senior Lecturer
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:
(Member)

Sifat Tanvir
Lecturer
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Dr. Md Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Ethics Statement

This research is conducted in strict adherence to ethical standards, utilizing publicly available and ethically sourced datasets to ensure respect for privacy and consent. The primary goal is to advance the detection of deepfake technology through visual artifact analysis and deep learning, aiming to safeguard digital media integrity and combat misinformation while upholding the highest standards of ethical responsibility in data usage and research practice.

Abstract

Our research introduces an advanced and robust framework in sentiment analysis for mental health detection on social media platforms using both standalone and hybrid deep learning models. Addressing the common challenges posed by noisy, informal, and semantically complex user generated content here in this research our study systematically investigates and compares the performance of various deep learning models such as : LSTM, RNN, MLP, and CNN architectures against a hybrid model that integrates the Universal Sentence Encoder (USE) with Gated Recurrent Units (GRU) deep learning model. By cleaning, and annotating a large-scale the multi-source dataset validated by our choosen psychologist expert review's our work demonstrates that conventional deep learning approaches often suffer from class imbalance and limited generalization, especially on real-world social data. On the other hand our hybrid (USE+GRU) model achieves a substantial improvement with an accuracy of 83% and significantly better precision-recall tradeoffs across sentiment classes, outperforming all baselines.These research findings decisively establish the superior potential of combining transfer learning with sequential modeling for nuanced sentiment detection and early mental health risk assessment in digital environments. The results not only extend the state of the art in text-based emotion and mental health analytics but also provide a scalable and more robust foundation for intelligent, ethical, and context aware intervention tools that can empower online platforms and researchers in monitoring well-being and mitigating the growing mental health crisis signaled through social media. Hence our research work sets a new benchmark for the practical deployment of deep learning in digital mental health and opens the way for the next generation of adaptive data focused solutions for social listening and psychological support in the near future.

Keywords: Sentiment Analysis, Depression detection, Mental Health Analysis, USE, Text Classification, Deep learning, F1, Recall, Precision, Hybrid.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis has been completed without any major interruption.

Secondly, to our Supervisor Tanzim Reza, Senior Lecturer sir for his kind support and advice in our work. He helped us whenever we needed help.

Thirdly, to our Co-supervisor Sifat Tanvir, Lecturer sir for his availabilty and support in our work.

And finally to our parents without their support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

Declaration i

Approval ii

Abstract iv

Acknowledgment vi

Table of Contents vii

List of Figures ix

List of Tables x

Nomenclature x

1 Introduction 1
1.1 Research Objective 2

2 Literature Review 3

3 Methodology 12

4 Dataset Preparation and Merging Techniques 15

5 Model Implementation Details: 18
5.1 LSTM: 18
5.2 RNN 20
5.3 MLP : 23
5.4 CNN: 25
5.5 Hybrid Model (USE + GRU) 27

6 Result and Analysis 29
6.1 LSTM Confusion Matrix: 29
6.2 RNN Confusion Matrix: 31
6.3 MLP Confusion Matrix: 32
6.4 CNN Confusion Matrix: 33

7 Conclusion 37

8 References 38

Bibliography **41**

List of Figures

3.1 Workflow 12

4.1 Flow chart of train-validation-test set 16
4.2 Data Split Diagram 17

5.1 LSTM Model architecture 19
5.2 RNN Model architecture 21
5.3 MLP Model architecture 23
5.4 CNN Model architecture 25
5.5 Hybrid (USE + GRU) Model architecture 28

6.1 LSTM Confusion Matrix 30
6.2 RNN confusion Matrix 31
6.3 MLP Confusion Matrix 32
6.4 CNN Confusion Matrix 33
6.5 Hybrid Confusion Matrix 34
6.6 Accurancy Comparison: Standalone vs Chosen Hybrid Models 35

List of Tables

6.1 Performance Metrics of Different Models for Sentiment Analysis . . . 36

Chapter 1

Introduction

Social media in the recent years has been a very effective way of expressing the self of individuals as people are able to express their ideas, feeling, and opinions concerning a broad range of matters in real time. Social media platforms, like Facebook, Instagram, X (previously known as Twitter) and Redditt's are highly useful tools because they produce large amounts of user-generated content on a daily basis, and thus can be used to study the general opinion of the audience. However, it is quite difficult to derive any meaningful insights out of this data that is unstructured, noisy, and of diverse character. Sentiment analysis as a sub-field of Natural Language Processing (NLP) is a problem in which subjective information is extracted and in turn used to determine the emotive tones used by the user in the expression. It is of immense importance in situations like market studies, customer opinions, political predictions, and monitoring of the opinion of the public. This study discusses a deep learning model of sentiment analysis and text classification on social media posts. Our preprocessing pipeline includes all necessary steps in order to clean the data and tokenize it, including solving the problems of slang, emojis, and misspellings currently common in social media language. Some types of deep learning models are used and tested in their ability to perform sentiment classification tasks, as Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN). Also, we suggest a hybrid model of using Universal Sentence Encoder (USE) and Gated Recurrent Units (GRU) to perform better in terms of classification accuracy and generalization results. The first aim of the current research is to evaluate the work of these models in terms of the correct classification of sentiments (positive or negative) based on social media data in order to reveal the feasibility of the advanced algorithms of natural language processing of the noisy and real-life textual data.

1.1 Research Objective

this research seeks to highlight the practical challenges involved in deploying deep learning solutions on raw social media content such as handling informal language, dealing with ambiguous context, and scaling models for large volumes of data. By comparing the performance of multiple architectures following means our study not only identifies the strengths and weaknesses of each approach but also provides insights valuation of which models are better suited for real time applications and mental health monitoring which is ultimately, the findings aim to inform future developments in intelligent systems capable of understanding human emotions and mental health signals in digital communication.

The insights gained from this analysis are intended to guide the selection and optimization of deep learning architectures for specific real world applications e.g. : real time sentiment analysis, social listening platforms, and automated mental health monitoring tools. These applications have far reaching implications, from enabling organizations to look into public sentiment and detect emerging trends to facilitating early intervention in mental health crises by identifying the risk individuals based on their online status or activities.

The research aims to :

1. Identify Challenges in multi social media sentiment analysis.

2. Evaluate multiple deep learning models

3.Optimize models for sentiment analysis

4. A suitable robust framework in better sentiment analysis.

By addressing both the theoretical and practical aspects of sentiment classification, this research aspires to advance the field of natural language processing and contribute meaningfully to the broader discourse on mental health analytics and digital wellbeing.

Chapter 2

Literature Review

According to the research paper [1], they collected data from social media sites, specifically Twitter, to determine emotion and sentiment using natural language processing (NLP). They assemble 10,000 tweets to train their model by dividing the data in an 80:20 ratio. Data collection and processing steps include app authentication, keyword list, text extraction from JSON, data cleaning, CSV creation, tokenization, stemming, stop word removal, and POS tagging. Specially, they have used Multinomial Naive Bayes Variant on account of its simplicity and accuracy and also Support Vector Machine(SVM) for linear and nonlinear classification.Here Multinomial Naive Bayse had an accuracy of 83 % and an F1 score of 83.29, whereas SVM had an accuracy of 79% and an F1 score of 79.73.

The paper[2] used multiple datasets to identify cognitive disorders. The used datasets are CowDist, MH-C, and MH-D. This paper used a Machine learning algorithm for automatic identification and categorization to detect mental health issues. In an online therapy dataset, their algorithm gave an F1 score of 0.88. To get this accuracy they perform unsupervised learning into the dataset. However, limitations in data collection methods are recognized, as well as difficulties in recognizing particular distortions because of short text passages and multiple distortions. The work emphasizes how crucial machine learning is to comprehending cognitive distortions and raises the possibility of noninvasive tools helping doctors spot and treat patients' erroneous thinking. With a particular focus on Reddit data, this paper[3] investigates the automatic identification and categorization of mental diseases from texts shared on social media in general. The study classifies nine mental health problems using cutting-edge deep learning models such as BERT, RoBERTa, and XLNET, showing better accuracy over earlier studies. The findings point to the possibility of automated techniques in early detection by showing that concentrating on broad text and post-level classification is sufficient for detecting mental diseases. BERT gained recall scores starting from 0.66 to 0.83 then F1 scores starting from 0.68 to 0.81, moving on precision 5 scores between 0.63 and 0.82. These findings show how well the classifiers identify a range of mental health issues.These findings show how well the classifiers identify a range of mental health issues. This paper gives an overview of significant factors to the field of mental health detection using natural language processing(NLP) techniques by addressing ethical issues.

At paper[3], Based upon the data found online, researchers have greatly explored

how posts and contents shared in social media affects mental health of people nowadays. At the beginning, data mainly from Twitter had been accumulated to aid the study. Various data collected from a number of sources have greatly helped in analysing the study, such as characterised patterns in tweets to figure out depression and PTSD by Coppersmith et al. (2015), different machine learning models by Benton et al.(2017) and signs of depression combining word embeddings with neural networks by Orabi et al.(2018). Switching from earlier sources such as twitter, researchers then shifted to more common mainstream used platforms such as Reddit that allowed longer posts to be shared with more details. Examples would include Gkotsis et al.(2017) using various models to detect mental health issues while Sekulic and strube(2019) focused on attention based networks to develop separate models for individual illnesses. In this paper Using a single RoBERTa-based model, researchers have been able to detect, classify and analyze multiple mental illnesses. The RoBERTa model outperformed both the BERT (F1 up to 0.87) and the LSTM models (F1 up to 0.76), achieving F1 scores of up to 0.89 on posts and titles. This demonstrates RoBERTa's ability to recognize the context of social media posts and provide more accurate findings across many mental health categories.

This paper[4], presents the information on the application of RoBERTa to the detection and classification of mental illnesses on social communication platforms. ADHD, depression, bipolar disorder, PTSD, anxiety and a non-mental illnesses's class, these five mental disorder forms have been recognized in this paper. The RoBERTa model behaves more effectively than the LSTM model at spotting posts that are irrelevant to mental illness. Behavioral tests, like synonym replacement and masking, have been utilized to evaluate the model's efficiency. The findings reveal that the model performs weakly in certain instances, indicating that the initial words must be contained in the text being used as input. In general, the investigation gives useful data for initiatives related to public health by displaying how effectively RoBERTa may recognise and categorize mental health illnesses on social media platforms.

The paper[5],Liu et al. (2018) erected a deep learning approach of the automatic detection of the presence of apple leaf diseases and convolutional neural networks (CNNs). It is a study where the researchers suggested an enhanced CNN architecture, that is the AlexNet, yet made to use less number of parameters, and yet is as efficient. Applied on a dataset consisting of more than 13000 apple leaf images, the model improved the classification accuracy to 97.62%. This work was an attempt to ascertain that deep convolutional neural networks are strong tools in feature extraction and disease classification on complex agricultural image datasets that make them useful in precision agriculture and smart farming.

The paper[6], explores data from Twitter and Reddit to identify mental issues of users using various machine learning algorithms. The paper mainly focused on using Natural language processing (NLP) and Machine Learning(ML) to analyze user data. In the initial stage of research, they apply character-level language models to create character sequences. Also identified a variety of mental health conditions using neural Single-Task Learning (STL), regression, and Multi-Task Learning (MTL) models. Later on, they employed CNN, GRU, Bi-GRU, and Bi-LSTM deep learning

algorithms. According to the paper's findings, Bi-LSTM was the best algorithm for identifying mental illness, with accuracy levels that were on par with those of BERT, a transfer learning model that has been trained beforehand. Among those algorithms, logistic regression and LinearSVC give an accuracy of 0.79 whereas others give lower accuracy. In this paper, they also include pre-trained transfer learning algorithms which give a 0.80 accuracy score. The importance of using ML, deep learning to improve the accuracy of identifying mental health has been shown in this research.

At Umarani [7], they have made an exploration paper on sentiment analysis on restaurant data utilizing different AI classifiers and deep learning techniques, for example, Multinomial Naive Bayes, Bernoulli Naive Bayes, Logistic Regression, SVM, Random Forest, K-nearest neighbor, Decision tree and LSTM, Convolution Neural Network. They have performed feature selection to reduce input factors and feature extraction to diminish computational complexity and to expand the model accuracy. They utilized AI strategies to change over text into vector structure and deep learning techniques used word embedding strategy to change over text into vector structure. The accuracy score of every classifier was tried by k-fold cross validation technique. After every one of the trials, the deep learning strategy CNN had the most elevated accuracy of 84.5 % though the machine learning classifier Decision Tree had the least 70%.

As per the paper[8], Twitter and IMDB are the main sources of data to analyze sentiment. The paper focused on Machine Learning Algorithms By way of illustration of Naive Bayes(NB), Support Vector Machines(SVM), and Artificial Neural Networking(ANN) to classify data. Their research is concentrated on using algorithms like TF-IDF to quantify team weights in documents and WordVec to constitute words in a vector space. The research uses three major approaches for classifying: Support Vector Machine (which has an accuracy of 83% on IMDB and 82% on Twitter), and Naive Bayes (which has an accuracy of 82% on IMDB and 72% on Twitter). Finally, 89% accuracy on IMDB and 86% accuracy on Twitter were attained using Artificial Neural Network.

In this paper[9] they have collected datasets to identify patient's disorders such as Schizophrenia, Autism, OCD, PTSD using patient's posts and their feedback from reddit. They have pre-processed the data using a variety of methods, including word stemming, tokenization, vectorization, and punctuation removal. The dataset was then tested using a variety of machine learning techniques, including Support Vector Machine, XGBoost, and Naive Bayes Classifier. For each component of the condition, the XGBoost classifier received the greatest accuracy score—58 This research[[11] is mainly based on machine learning and text classification to analyze mental health. In this paper they used text mining technologies meanwhile semantic information is removed to detect mental health issues. To diagnose mental health issues into proper categories they perform multiple trials. Support vector machines (SVMs) are used for text categorization tasks, with high accuracy (92 At [5], they gathered 10,000 datasets from mental health-related subreddits (BPD, Bipolar, Depression, Anxiety). The dataset was prepared using a variety of data pre-processing techniques. They eliminated the urls first to avoid any web links that could interfere

with the study. They also deleted the punctuation marks to highlight the textual content. Finally, they converted raw text documents into a TF-IDF feature matrix, which measures each word's significance in the texts according to its frequency and rarity over the corpus. Then, they used three machine learning algorithms: Light GBM, Multilayer Perception, and Naive Bayes Classifier. Ultimately, the multilayer perception had the lowest accuracy score of 71.4%, while the Light GBM classifier obtained the highest accuracy score of 77%.

At Paper[10], The review tells us that, to detect sentiment through machine learning many researchers have used social media platforms like twitter and Facebook. These studies show users posts to find emotional and language patters linked to sentiment. In terms of better accuracy common algorithms are used which are Naïve Bayes, Support Vector Machine (SVM), Decision Trees and a hybrid model. The best result in terms of accuracy has been showed by Naïve Bayes. In some of the studies, hybrid model which includes Naïve bayes combined with SVM has also been used which showed good results in sentiment analysis. Many of the previous studies had limitations and most of them focuses on just one ml model and single dataset which is also limiting the results. The main objective of this paper is to improve the past efforts by comparing two different algorithms which are Naïve Bayes and NBTree. The dataset is also larger and is collected from twitter. The goal is to find a more accurate and reliable method to detect accurate sentiment based on the tweets.

In this paper[11], Many present studies have utilized machine learning to evaluate mental health through publicly available social media data. Prior research utilized methods that mainly included SVM and Naive Bayes for sorting mental health-related texts and sentences. The analyses mainly demonstrated that blending multiple sources of data (especially forum discussions and inquiries) substantially better outcomes. Deep learning tools like Hierarchical Attention Networks mainly raised precision by recognizing the context in user inputs. Several research projects also used combined models for example Random Forest and XGBoost. They performed effectively but needed balanced as well as high-quality datasets. The majority of studies conclude that machine learning is helpful for detecting mental health. But efficiency is determined by the dataset-data preprocessing as well as the models that have been selected. The present study mainly demonstrated that Random Forest had the highest accuracy (97%) while evaluating social media material as indicators of sentiment analysis.

At paper[12], Many researchers have tried to detect mental health issues such as stress and anxiety by analyzing social media posts. Previously, research was conducted on surveys , voice data or on small groups of people, limiting the generalizability of the findings. These studies were also challenged with tiny datasets, biased reporting, and problems interpreting the emotional content of the words individuals used. Older methods frequently depended on basic machine learning and lacked a thorough understanding of language context. This makes it difficult to distinguish complicated emotions, like anxiety, stress etc. Recent methods include Natural Language Processing (NLP) models such as decision trees, random forests, and, most notably, BERT, which can interpret language and emotions better than before. Employing a vast number of social media datasets and progressed models has enhanced

the accuracy and reliability of mental health predictions. According to the content they put online these tools give a better way to understand people's mental states.

At paper[13], Various studies of the social media sentiment analysis based on Twitter data have identified some major approaches and conclusions. Support Vector Machines (SVM) and Naive Bayes are the most commonly used machine learning techniques due to their simplicity and their high level of efficiency, however, these techniques may not work with complex language patterns and context. The recurrent Neural Networks (RNNs), especially the Long Short-Term Memory (LSTM) networks, provide an improved contextual understanding and have a higher accuracy, but with increased computer resource requirements. Approaches deploying a lexicon, e.g. with VADER or SentiWordNet, assign a sentiment score to a word and are appreciated since they are fast, simple, but usually do not recognize sarcasm or subtle language. Ensemble models by adding several classifiers have gained popularity to enhance performance at the expense of complexity and increased model computational needs. This work has used sentiment analysis performed at a sentence level to identify finer emotional expressions. Such feature extraction strategies as Bag-of-Words, word embeddings (such as Word2Vec), n-grams, emoji detection play a critical role in transforming a textual data into machine-readable routines. Through performance measurements, it can be demonstrated that the best accuracy of 86% is attained by Linear SVC, and the second best is obtained by Logistic Regression and Random Forest. Nevertheless, quandaries that face such approaches are still significant, including language ambiguity, algorithmic bias, and privacy concerns. The future work, to overcome these problems, will be to continuously improve the models, to be more ethical in the processing of the data, and be more precise in translating the sentiment of the people.

At paper[14],In this study the researcher detected depression using a variety of artificial intelligent methods in Arabic text. They gathered almost 19000 posts and selected 2500 depressed posts and 2500 non-depressed posts from the Arabic mental health forum "Nafsany" to create a balanced dataset. They employed three methods: a lexicon-based method, deep learning (DL) methods, and machine learning (ML) models such as Logistic Regression, Naive Bayes, SVM, Random Forest, etc. Deep learning-wise, they developed an ensemble model integrating AraBERT with layers of LSTM, BiLSTM, and GRU. The combined deep learning model outperformed all other models, reaching 93.03% accuracy and a 92.93% F1-score.By contrast, the lexicon-based model achieved just 73.31% accuracy, while the best machine learning model (SVM with TF-IDF) attained approximately 88.72% accuracy. For detecting depression in Arabic literature, the hybrid deep learning ensemble was throughout the most accurate and dependable.

At paper[15], This study proposes a new way to read tweets by mixing BERT with LSTM so machines grasp sentiment better. Working with 1.6 million labelled tweets from Sentiment140, the team let BERT build a rich word sequence and used LSTM to track their order.The purpose of this combo was to enhance the significance of tweet sentences. In a series of tests, the researchers compared their hybrid BERT-LSTM to self-contained deep learning networks and conventional machine-learning techniques. The combination exceeded all benchmarks on accuracy, precision, recall,

and F1, however specific figures are not disclosed. Additionally, they point out that compared to previous approaches, their system produced labels that were clearer and more reliable. As a result, the study finds that the BERT-LSTM combination was the most effective model it explored.

At paper[16], This study depending on sentiment and emotion recognition presents a model for positive, negative, and neutral emotion detection in social media messages combining Punjabi and English. CountVectorizer was utilized for cleaning and vectorizing the dataset of 10,307 testing and 37,805 training posts. For the actual classification of emotions, authors used a Logistic Regression simple model. Regular evaluation metrics; accuracy, precision, recall, and F1-score were utilized to judge performance. Logistic regression model demonstrated a very high accuracy of 92%, with precision and recall perfectly around 0.92 for all classes. This single-model approach benefits code-mixed language sentiment analysis as the model performs so well and uniformly across all emotion classes. Therefore, Logistic Regression turned out to be both the top-performing and main model in this research.

At paper[17], This research investigates the capacity of multiple deep learning models to detect hate speech on social media platforms. The models in study included five popular models: LSTM, BiLSTM, GRU, CONV1D, and BERT. A balanced and clean dataset of social media posts-for which all special characters and extraneous materials were stripped-was used to train each of the models. The performance of the models was evaluated by means of F1 score, recall, accuracy, and precision. The BiLSTM model obtained the highest accuracy of 99.37% and an F1 score of 0.99 and hence proved to be very competent at differentiating between hate speech andnon-hate speech confidently. Likewise, GRU and LSTM performed the other two best, with an accuracy of 99.34% and 99.14%. The remaining mutators, CONV1D and BERT, took the next places with 99.24% and 99.04% accuracy, respectively. BiLSTM became the best choice for hate speech recognition in this study because it gave the best balance between precision and recall. The study emphasizes that although these models are effective, their capacity to handle more complex and varied hate speech may be enhanced in the future by utilizing larger and more varied datasets.

At paper[18],This research focuses on using deep learning methods to perform sentiment analysis over social media texts used in informal language and slang, particularly the Indonesian language.Recherchars worked with three fundamental deep learning models which are: Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNNs). Social media text tends to be full of abbreviations, slang, and non-standard grammar, and thus using conventional machine learning techniques does not quite work with it. To fix this problem, the authors utilized several text preprocessing procedures such as text normalization, slang lexicons, and tokenization to clean and ready the data. The research evaluated the models' performance using accuracy, precision, recall, and F1-score after training and modeling them on social media posts from Facebook, Instagram, and Twitter. Compared to any other model, LSTM was best in long sequence texts and also in-depth determination context over RNN and CNN. Therefore, to classify sentiment in categories like positive, negative, and neutral the

LSTM model is more reliable than other models. It was also stated in the study that text normalization was important in acting as a precondition to enhance model performance. In conclusion, the paper demonstrates the efficacy of deep learning (more especially, LSTM) in the task of sentiment analysis on social media texts that contain slang and informal language because LSTM can handle linguistic complexity and long-distance reliance on word position.

At paper[19],The present paper entails in-depth research on the application of the Natural Language Processing (NLP) technique of sentiment analysis in the social media marketing practice. It is concerned with studying the opinions and sentiments of a post made by people about the brands or the products. The researchers intend to use a hybrid model that integrates the advantages of deep learning methods like CNNs and RNNs with reinforcement learning approaches like Q-learning and SARSA. Q-learning was used to optimize the model, which was to minimize the deep learning model in terms of feedback. The data were gathered from social media like Facebook, Instagram, and Twitter. It underwent many preprocessing stages, involving converting text to lowercase, eliminating URLs, stop words, and special characters, and cleaning and preparing it for analysis. Using TF-IDF and Bag-of-words methods the data were collected and sent to a model of neural network. To test this model a survey was conducted and this survey showed success identifying people's diverse feelings towards the product. The end model provides accuracy of 96% in sorting data into three categories positive, neutral, and negative. The study found that 70% of people had a positive opinion towards the product where 20% had neutral and only 10 % had a negative opinion on the product. The results make it obvious that pairing CNN, Q-learning and RNN gives higher reliability analyzing opinions of consumers.

At paper[20],The research made a comparison between multiple machine learning algorithms to analyse emotion and sentiment. To analyse the sentiment they used online data and collect it from Amazon product reviews and Twitter posts. For the tasting phase the algorithms used are Logistic Regression (LR), Random Forest (RF), and Naive Bayes (NB). Before the testing phase those data were cleaned and prepared by tokenization, removing common words and punctuation marks and stemming. afterwards the text was converted to numbers using vector methods so that the models could process it.Precision, recall, accuracy, and F1-score were used to evaluate each model's performance in identifying the text's sentiment. Naive Bayes outperformed all other models with 93.9 accuracy, 95 precision, 98 recall, and an F1-score of 97%, making it the best model for structured and balanced review data sets. Random Forest has the highest accuracy rating on the Twitter sentiment list at 94%, indicating the resilience of social media data. Overall, the data appears to demonstrate that Naive Bayes is best suited for examining formal product reviews, but Random Forest excels at analyzing social media posts. This research highlights the variations in algorithm selection with respect to the type of data and suggests that, in order to have a better knowledge of how the model performs during the sentiment prediction task, numerous evaluation measures should be used.

At paper[21], A review of machine learning application to sentiment analysis in social media and e-commerce was made by Nawrocka et al. (2025). Conventional

approaches, such as the use of lexicon-based and old fashioned machine learning models (e.g. SVM, Decision Trees) proved useful but far too weak to work with difficult language. The deep learning models, in particular, LSTM outperformed due to their ability to capture the context and the sequence data. They ran their model on Twitter and Reddit data and succeeded at an accuracy of 92 percent using LSTM with the activation function of the tangent and an input length of 1000. They implied that there was more to come in transfer learning as well as model optimization.

At paper[22],This essay will dwell on measures of sentiment analysis in different social media. It brings out the use of machine learning, deep learning, and hybrid models in classifying the sentiments of the users. The most important ones are data preprocessing, feature extraction, and model optimization. The paper examines approaches such as SVM, Naive Bayes, CNN, and LSTM and gives an overview of sentiment analysis categories, i.e., emotion and aspect-based analysis. It is focused on the importance of using social media feedback as a source of business decision-making.

At paper[23], Big data analytics in social media is the subject of the systematic review in this paper. It proposes the Sunflower Model, which expands 5Vs of big data to five more socio-technical dimensions. There are ten key types of analytics and text analytics is the most popular. The study also provides a taxonomy and enumerates machine learning methods applied to each of the different types of analytic to enable researchers to choose adequate methods to be used in analyzing various social media data.

At paper[24], This research paper gives a sentiment analysis approach that uses Naive Bayes to categorize user comments into positive, negative, neutral, or irrelevant categories from social media. The system does sentiment scoring, text and emoji analysis, and automatic tweet and comment processing. It is intended to enable a user to quickly understand the opinions of a crowd without having to read through every comment.Measures including precision, recall, and F1-score were employed in the evaluation procedure, and the accuracy of the provided model was 98%. This can process any volume of massive social media data and will be more effective, efficient, and faster than the manual mode or other earlier models. By encouraging people to apply emoji-based analysis, which adds depth to the emotion recognition and helps the user make better selections, the system bases its decisions on online feedback.

At paper[25], The research analyzes the Naive Bayes algorithm's capacity to detect the existence or lack of Pancasila spirit, Indonesia's defining ideology, as expressed in tweets produced by the tool's social media users. The tweets are automatically categorized by the system into three groups based on whether they are neutral, negative, or positive. A collection of data of 100 Twitter accounts, each with a maximum of 100 tweets, was used and preprocessed using tokenization, stemming, and stop word elimination. The model had a classification accuracy of 73% in both training and testing. It also demonstrated great precision (up to 94%) and recall (up to 90%), particularly when detecting negative sentiments. Despite the fact that the model was not particularly helpful in identifying neutral content, the research shows that

Naive Bayes is a viable and effective technique for investigating ethical behavior and national values on social media. Such an approach offers the possibility of tracking digital conversation and supporting value-driven education and policymaking.

At paper[26], The article Social Media Text Sentiment Analysis: Exploration of Machine Learning Methods studies how three types of machine learning algorithms, namely logistic regression and recurrent neural networks (RNN), perform to categorize the sentiment of Amazon Kindle books based on the reviews of the books. The post-processing of data cleansed helped the researcher compare the specificities of model performance and find that logistic regression proved more accurate (82.2%) when compared with RNN (79.25%) even though the latter is better at working with contextual undertones. Analysis of the feature weight showed that there are determinants of emotion within the text, which can be interpreted. Real-time sentiment analysis and visualization was also done via an equally developed prototype web app. The research underlines the complementary nature of traditional and deep learning frameworks with regard to sentiment classification and recommends hybrid modeling in future studies.

Chapter 3

Methodology

Our methodology followed in this research involves a very well suited structured and systematic approach to classify mental health sentiment analysis using textual data in CSV format from various social media platforms. The whole process is clearly described in several interconnected stages such as : data collection, preprocessing, feature extraction, model construction, and evaluation.

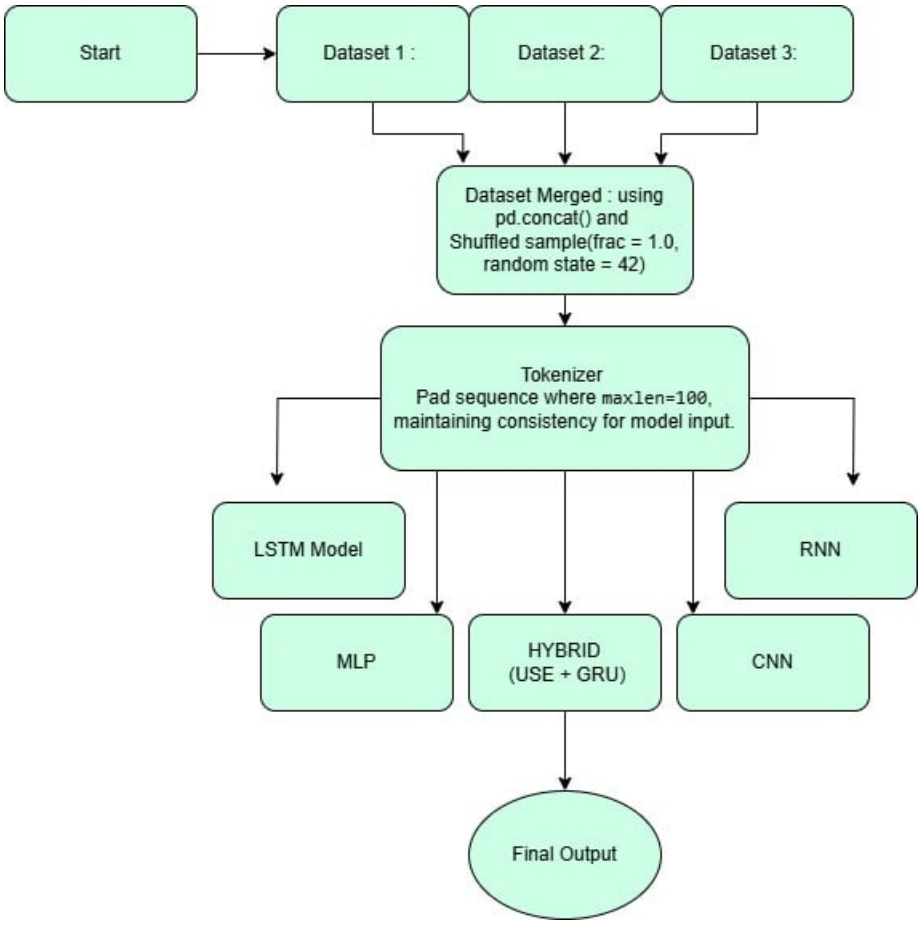


Figure 3.1: Workflow

1. Data Collection and Integration: The Collected Three distinct datasets related to mental health posts from social media were initially choosen. To enhance model performance through a diversified dataset, these were merged into a single entity dataset.The merging was performed using pandas concat() method. Post-merging the dataset was shuffled using sample frac=1.0, random state=42 to eliminate any biases that may come from the original ordering of data ensuring random distribution across the combination of all dataset into a final dataset.

2. Text Preprocessing and Tokenization: The textual data underwent extensive preprocessing including cleaning , normalization and conversion into a machine readable format. The texts were tokenized into sequences and to maintain a consistent input size suitable for neural network models following these sequences were padded using a maximum length where we used maxlen=100. This padding step ensures uniformity across inputs, facilitating efficient model training and consistent output.

3. Model Construction: The research explored multiple modeling approaches that is aiming to evaluate and compare their effectiveness in mental health sentiment analysis. We have chosen four Deep Learning (DL) models along with our own hybrid model. Long Short-Term Memory (LSTM): A recurrent neural network variant adept at capturing long term dependencies in sequential data. Multi-layer Perceptron (MLP): A fully connected neural network used here for classification tasks offering baseline performance. Recurrent Neural Network (RNN): A basic sequence-modeling neural network tested to benchmark performance relative to LSTM. Convolutional Neural Network (CNN): A network traditionally used in image processing but applied here to text data, utilizing its capacity to detect local patterns and important features. Hybrid Model (USE + GRU): This advanced hybrid approach combines Google's Universal Sentence Encoder (USE), known for producing semantic-rich embeddings, with a Gated Recurrent Unit (GRU), known for effective sequential data processing. The hybrid model aims to take advantage of the strengths of transfer learning (using USE) and sequential modeling (GRU).

4. Model Training and Evaluation: Each model was individually trained on the processed data set with clearly defined hyperparameters, including batch size, epochs, and learning rates. The models were evaluated based on standard metrics such as accuracy, precision, recall, and F1 score, with a focus on comparing standalone models against the hybrid approach.

5. Final Prediction and Output: The predictions of each trained model were thoroughly analyzed and compared. The hybrid model (USE + GRU) emerged with a superior performance profile, leveraging deep semantic embeddings of USE combined with the effective sequential capturing abilities of GRU. The final predictions of the best performing model are documented and discussed,

highlighting its applicability in accurately classifying mental health conditions from social media posts.

Chapter 4

Dataset Preparation and Merging Techniques

The dataset used in this research combines three distinct CSV files: the primary mental_health_sentiment_dataset.csv, the secondary sentiment_analysis.csv, and an additional dataset named Final Data.csv.The dataset used in this research combines three distinct CSV files: the primary mental health sentiment dataset.csv. the secondary sentiment analysis.csv and an additional dataset named Final Data.csv. We have collected 2 authentic datasets from Kaggle. The ratings of those datasets were 10 out of 10. Furthermore, both of them were MIT licensed dataset. So we made sure that the datasets were valid datasets. Also, we have collected around 500 raw data from social media such as Facebook, Instagram, X, Reddit. Then we have labeled all the data based on the post type(Positive or Negative). And we took approval from an experienced Psychiatrist named Dr.Jurdi Adam MBBS, MD (Psychiatry) BMDC Reg.No:A-65423 Psychiatrist of LifeSpring to validate our collected datasets. Initially, inconsistencies in column naming and sentiment labeling were resolved by standardizing column headers specifically, renaming the sentence column to text. Subsequently, all datasets were filtered to retain only the essential columns (text and sentiment), removing the rows containing missing data to maintain data integrity. To unify sentiment labels, a normalization function converted diverse sentiment indicators, such as "positive", "neg", "1", and "0", into a consistent binary format, assigning 1 for positive sentiments and 0 for negative sentiments. Post-normalization, the datasets were concatenated into a single DataFrame and thoroughly shuffled to ensure randomness and eliminate any source-specific bias. The finalized dataset comprises 28,016 samples with a distribution of 17,841 positive and 10,175 negative entries, thereby providing a substantial and balanced corpus suitable for robust model training and evaluation.

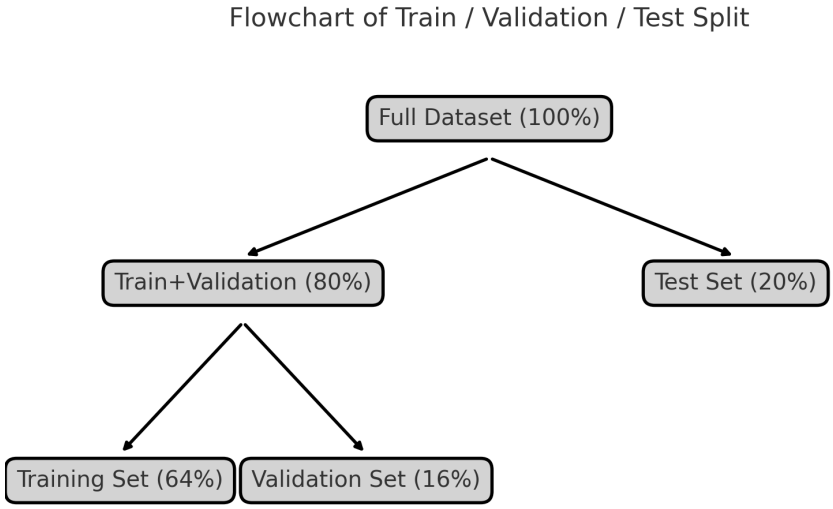


Figure 4.1: Flow chart of train-validation-test set

We used, merging and standardizing multiple data sources with normalized sentiment labels the consolidated dataset undergoes a series of pre processing steps designed to prepare it for robust model training and evaluation. Initially the text data is tokenized and padded to create uniform input sequences a critical step ensuring that inputs are in a consistent format for downstream neural network architectures. The dataset is then split into training and testing subsets using an 80/20 ratio via the train_test_split function, where the stratify=y parameter is employed to maintain the inherent class imbalance thereby preserving the original distribution of positive and negative classes in both subsets. Following approach of splitting not only facilitates unbiased evaluation of model performance but also reinforces the stability and generalizability of the resulting sentiment analysis model in the context of mental health detection.Additionally we used, during the training phase some portion of the training set is set aside as a validation set typically through the validation_split parameter in model training routines to monitor performance and facilitate hyperparameter tuning. The validation set plays a crucial role in implementing early stopping, where training is halted if the model's performance on unseen validation data fails to improve, thus preventing overfitting and enhancing the model's generalizability in detecting nuanced mental health sentiments.

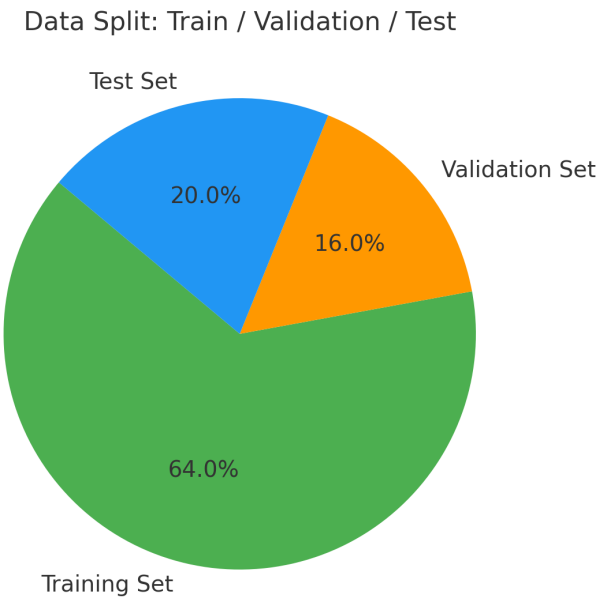


Figure 4.2: Data Split Diagram

Finally now, after the following merging and preprocessing the datasets the final data was tokenized and padded to create uniform input sequences for the model. The dataset was initially split into an 80/20 ratio using a stratified approach via the train_test_split function, ensuring that the distribution of positive and negative sentiments remained consistent across subsets. From the 80% designated for training further partitioning was performed allocating 20% of this portion for validation which ultimately made our final data that is : resulting in a final split of 64% for training, 16% for validation, and 20% for testing. This three level of splitting strategy allows for robust model training with continuous hyperparameter tuning based on validation performance, while the test set remains a completely unseen subset to reliably assess the model's generalizability.

Chapter 5

Model Implementation Details:

The research explores various neural network architectures, including Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Simple Recurrent Neural Networks (RNN), Multilayer Perceptron (MLP) and finally our proposed hybrid model combining the Universal Sentence Encoder (USE) with GRU chosen strategically to capture the nuanced patterns of mental health sentiment in text. LSTMs excel in preserving long range dependencies, making them ideal for detecting complex emotional expressions while on the other hand GRUs offer a streamlined alternative with fewer parameters and faster computation, crucial for real-time applications. RNNs, despite their vanishing gradient limitations, serve as a baseline to evaluate the effectiveness of more sophisticated architectures, while MLPs assess whether sentiment classification is feasible using purely feedforward networks. The USE + GRU hybrid model strengthens semantic understanding by leveraging USE's contextual embeddings alongside GRU deep learning model's seaming alternate with faster computation for almost every real time application. Collectively,these models provide a holistic perspective on deep learning approaches for mental health sentiment analysis, ensuring robustness.

5.1 LSTM:

The LSTM model was chosen for its exceptional ability to capture long range dependencies in sequential data which is indispensable when dealing with text classification tasks where context and word order are pivotal. In this implementation the text sequences are first transformed into dense and fixed length representations using an embedding layer configured with a vocabulary size of 10,000 and an embedding dimension of 64, ensuring that each word is mapped to a continuous vector space over an input length of 100 tokens. This transformation facilitates the extraction of semantic features before the data are processed by a single LSTM layer comprising 64 memory units the parameter return_sequences is set to false to ensure that only the final hidden state which serves as a distilled representation of the entire sequence are forwarded to subsequent layers.
To further refine the learned features and introduce non linearity, the output of the LSTM is passed to a dense layer with 32 units implementing the ReLU activation function. this layer is augmented with an L1 and L2 regularizer (using coefficients of 1e-5 and 1e-4 respectively) to penalize large weights and thus mitigate overfitting, effectively acting as an implicit dropout by constraining the model complexity. The

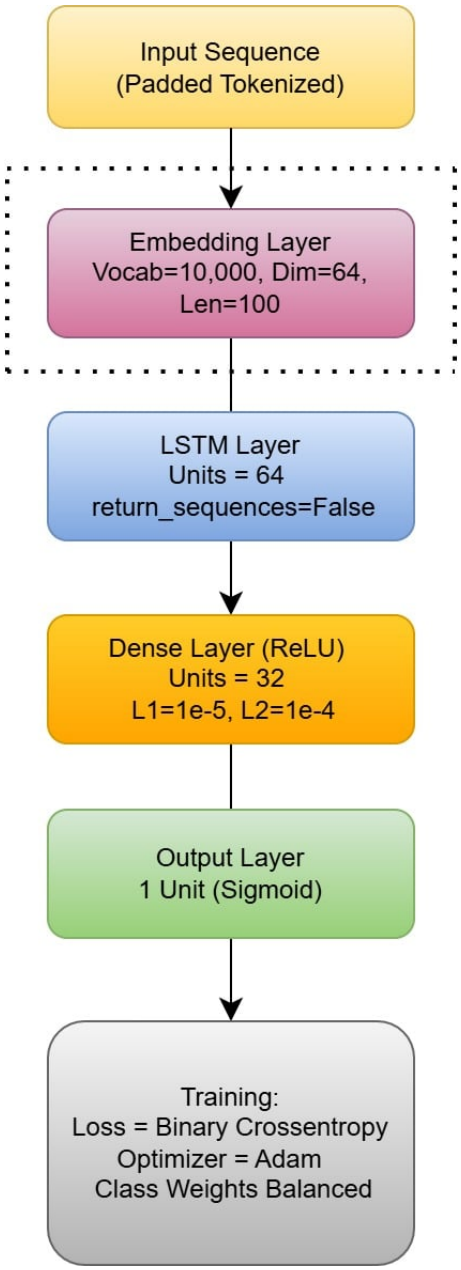


Figure 5.1: LSTM Model architecture

network concludes with a single neuron dense layer that uses a sigmoid activation function to output a probability score between 0 and 1, which is particularly suited for binary sentiment classification. Recognizing the imbalance present within the training data class weights are computed using a balanced strategy, ensuring that the minority class is given proportionally higher importance during the training process which is crucial for teaching the model to discern subtle signals in underrepresented classes. The model is compiled with the binary cross-entropy loss function, which is appropriate for binary classification tasks and the Adam optimizer is chosen for its adaptive learning rate strategy that accelerates convergence while maintaining training stability. Training is conducted over a maximum of 20 epochs with a batch size of 256 to efficiently handle large volumes of data here it is very importantly an early stopping mechanism is employed that monitors the validation loss and halts the training process if no improvement is observed over 3 consecutive epochs or parents value thereby preventing overfitting and ensuring that the model preserves its best learned weights for optimal generalization. This comprehensive configuration of the LSTM model with its robust embedding layer, carefully tuned recurrent architecture, regularized dense layer, and practical training strategies offers an smart solution for capturing the refinement context-dependent features of text data in a computationally efficient manner while rigorously addressing issues of class imbalance and overfitting throughout the training process.

5.2 RNN

Recurrent Neural Network (RNN) model implemented in our experiment is designed to efficiently capture the sequential nature of text data leveraging its ability to maintain an internal state that reflects previous inputs. Following this makes it particularly suitable for capturing temporal and contextual dependencies in sentiment analysis tasks. In this model the text sequences are first transformed into dense vector representations using an embedding layer with a vocabulary size of 10,000 and an embedding dimension of 64 which ensures that each sequence of maximum length 100 is projected into a continuous vector space where semantic relationships between words can be effectively learned. Following the embedding layer, the model employs a SimpleRNN layer with 64 hidden units this layer processes the sequential data by iteratively updating its hidden state in which encapsulates information from previous time steps so ultimately enabling the RNN to grasp the flow and context of the text. Although simpler in structure compared to advanced architectures like LSTM which include gates to mitigate issues like vanishing gradients the use of a SimpleRNN offers a computationally efficient alternative for tasks where elaborate memory mechanisms may not be critical, especially when computational resources or training time are considerations. The output from the RNN is then fed into a dense layer with 32 units activated by the ReLU function. this fully connected layer serves to further refine and abstract the features extracted by the recurrent layer by applying non linear transformations that improve the model's ability to discern subtle patterns.
Finally, the model concludes with a single-neuron dense layer using a sigmoid activation function, which outputs a probability score between 0 and 1 making it ideal for binary classification tasks such as sentiment detection. The entire model is com-

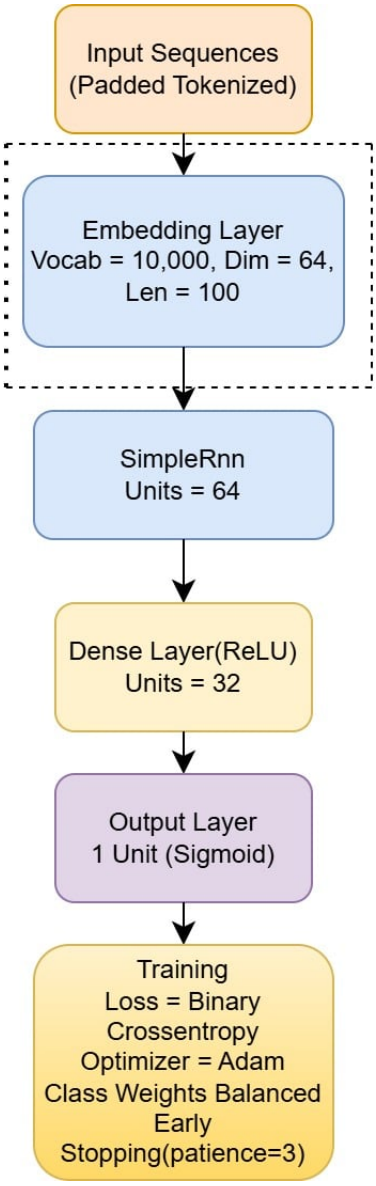


Figure 5.2: RNN Model architecture

piled with the binary cross-entropy loss function appropriate for binary outcomes, and optimized using the Adam optimizer, which automatically adapts the learning rate and enhances convergence stability during training. The network is trained over 20 epochs with a validation split of 20%, which allows for continuous performance monitoring on unseen data and helps to guards against overfitting issues. this following training patern is chosen to balance between adequate learning iterations and computational efficiency. In summary, the choosen SimpleRNN based architecture is strategically chosen for its simplicity and its inherent capability to process sequential data, providing a streamlined model that effectively captures the dynamics of language while remaining computationally lightweight and easier to implement relative to more complex recurrent architectures.

5.3 MLP :

In our research the Multilayer Perceptron (MLP) model implemented is a feedforward neural network designed to perform binary sentiment classification by transforming preprocessed numerical text data into a binary decision. The implemented Multilayer Perceptron (MLP) model is structured as a sequential feedforward network tailored for binary classification tasks. The model begins with an input layer accepting a 100 dime vector which represents the preprocessed text data. This is passed to a dense layer with 128 neurons utilizing the ReLU activation function to capture complex non linear patterns in the features. A subsequent dense layer with 64 neurons, also using ReLU, further abstracts these learned representations. Finally a single-neuron dense layer with a sigmoid activation function outputs a probability value between 0 or 1 making it ideal for binary sentiment classification.

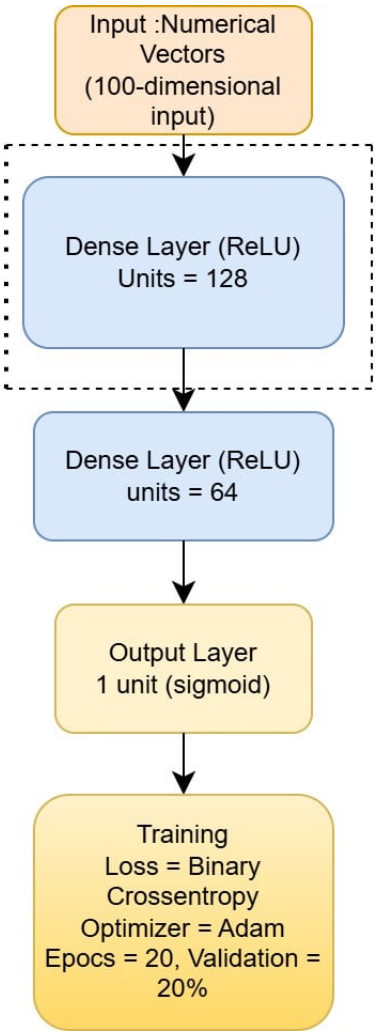


Figure 5.3: MLP Model architecture

The model employs the binary cross entropy loss function and is optimized using the Adam optimizer, renowned for its adaptive learning rate that enhances convergence speed and stability. Training is executed epoch = 20 which is a choice balancing sufficient learning iterations with the risk of overfitting along with a validation split of 20% from the training data to monitor generalization performance during training. This architecture was chosen for its simplicity and effectiveness in modeling non-linear relationships, providing a computationally efficient baseline that can be further fine tuned or expanded upon for more complex applications.

5.4 CNN:

We choosed The Convolutional Neural Network (CNN) model was for its remarkable ability to automatically learn local and spatial features from sequential text data, making it an appealing alternative to recurrent models for certain text classification tasks. In this implementation the model begins with a one-dimensional convolutional layer (Conv1D) that employs 128 filters with a kernel size of 3 and the ReLU activation function this layer scans through the 100-token input sequence (reshaped to include a singleton channel dimension) to extract meaningful local patterns or n-grams, capturing subtle linguistic features that may be indicative of sentiment regardless of their position in the sequence.

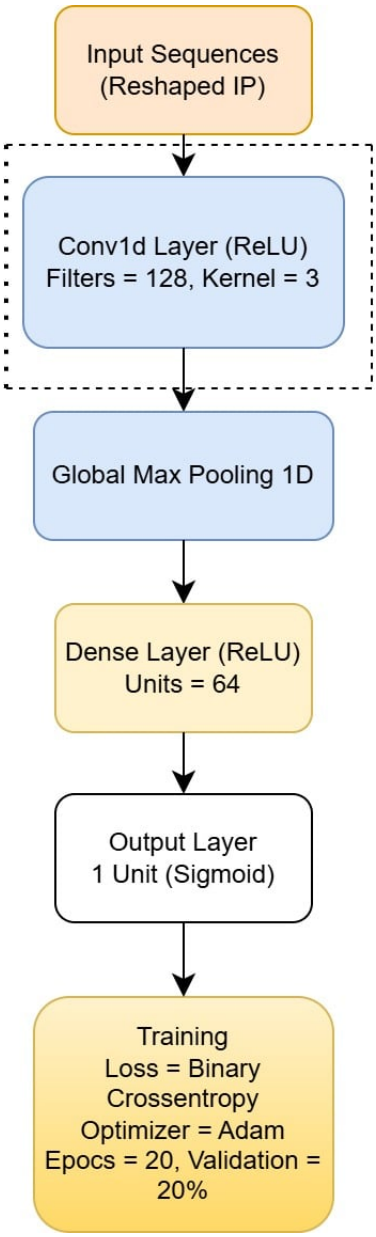


Figure 5.4: CNN Model architecture

Following the convolution, the architecture utilizes a GlobalMaxPooling1D layer which condenses the feature maps by selecting the maximum value from each filter's output the this approach not only greatly reduces the dimensionality of the learned representation but also introduces a degree of translation invariance, ensuring that the most features are collected irrespective of their data within the text. The pooled features are then fed into a fully connected (Dense) hidden layer consisting of 64 neurons with ReLU activation this layer further abstracts and refines the extracted features by combining them in a non linear manner to create a higher level representation conducive to classification. The final layer is a single neuron with a sigmoid activation function that outputs a probability between 0 or 1, directly mapping the learned features to a binary sentiment classification decision. For model optimization, the network is compiled using the binary cross-entropy loss function and optimized with Adam, which is very well known method for its adaptive learning rate that facilitates efficient convergence across varied data distributions, while accuracy is tracked as the primary performance metric. Training is carried out over 20 epochs with 20% of the training data reserved for validation, allowing continuous performance monitoring and ensuring that the model generalizes well on unseen data. This CNN architecture was selected because it offers a robust blend of computational efficiency and effective feature extraction and its ability to distill complex textual patterns into concise and informative representations makes it particularly suitable for tasks where rapid and reliable classification is essential without the added computational overhead of deeper or more recurrent-based models.

5.5 Hybrid Model (USE + GRU)

The hybrid model was selected to take advantage of the strengths of both sequential encoding and transfer learning, providing a more robust and semantically enriched representation of text for binary classification. In the first part of the implementation, class imbalance is addressed by computing class weights using sklearn's compute class weight, which ensures that underrepresented classes receive proportionally more emphasis during training. The model is then built using TensorFlow Keras's Sequential API, starting with an embedding layer that maps a fixed vocabulary of 10,000 words into 64-dimensional vectors for each token in sequences of length 100. This layer serves as the fundamental feature extractor that translates discrete word indices into continuous feature representations. GRU (Gated Recurrent Unit) layer with 128 units follows, set to output only the final state which ultimately captures the contextual information from the input sequence while remaining computationally efficient compared to more complex recurrent architectures.

For further refine the learned features and mitigate overfitting, a dense layer with 64 neurons and a ReLU activation function is incorporated and the dense layers are regularized using combined $L_1(1 \times 10^{-5})$ and $L_2(1 \times 10^{-4})$ penalties to constrain the weights, thus encouraging simpler and more generalizable models. The final layer is a single neuron with a sigmoid activation function that outputs a probability score, mapping the processed features into a binary decision. The model is compiled with the binary cross-entropy loss function and the Adam optimizer, which has an adaptive learning rate that helps in achieving faster convergence and stable training dynamics. Training is performed over 20 epochs with a validation split of 20% and a batch size of 256, while class weights ensure that the model remains sensitive to both positive and negative classes a key consideration when dealing with imbalanced datasets. In the second part of the hybrid approach, transfer learning is incorporated by loading the Universal Sentence Encoder (USE) from TensorFlow Hub, which generates semantically meaningful 512-dimensional embeddings for the raw text. These embeddings capture broader contextual and semantic nuances beyond mere word counts or sequences. The USE embeddings are reshaped to match the expected input dimensions for the GRU-based model and a new Sequential model is defined that starts with an input layer compatible with these embeddings. This model uses a GRU layer with 128 units to process the sequential structure embedded in the pre-trained representations, followed by a dense layer with 64 neurons activated by ReLU to further refine the features. The output layer remains a single neuron with a sigmoid function to produce a binary classification. Like its GRU-only counterpart, this USE+GRU hybrid model is compiled with binary cross-entropy loss and optimized using Adam, then trained over 20 epochs with a 20% validation split and an appropriate batch size = 128. This dual-component approach of combining traditional recurrent neural architectures with pre-trained sentence embeddings allows the model to benefit from both explicitly learned sequential patterns and the contextual richness provided by transfer learning, ultimately leading to improved classification performance while addressing issues like class imbalance and overfitting through regularization and appropriate training strategies.

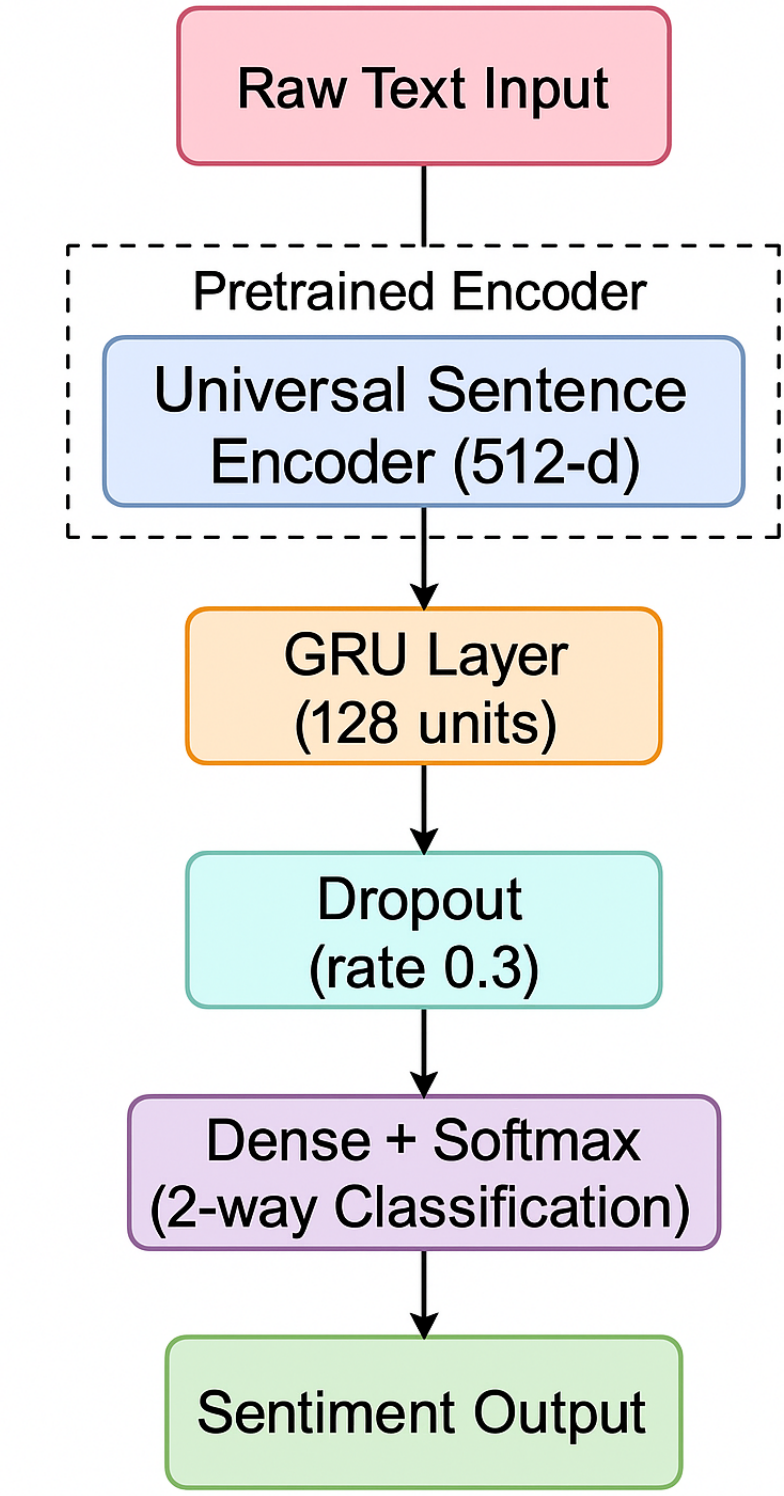


Figure 5.5: Hybrid (USE + GRU) Model architecture

Chapter 6

Result and Analysis

Below is the result and analysis part of our research. For the confusion matrix :

TP → Model correctly predicted Yes.

TN → Model correctly predicted No.

FP → Model wrongly predicted Yes (should be No).

FN → Model wrongly predicted No (should be Yes).

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Perfect accuracy would have high TP and TN, low FP and FN.

6.1 LSTM Confusion Matrix:

The LSTM achieved an overall accuracy of 64% but showed a pronounced bias toward the positive class: out of 2,035 true negatives only 13 were correctly identified 0.6% recall on negative, whereas 3,572 true positives yielded a recall of 99% on positive. This imbalance stems from the LSTM's strength in modeling long-range dependencies capturing sentiment signals that often span sentences but its tendency to overfit the majority class when the batch size = 128 units and dropout rate = 0.2 are not sufficiently tuned for severe class imbalance. In our noisy, informal social-media corpus with misspellings, slang, and sarcasm the LSTM's gates can latch onto frequent positive cues, flooding its hidden state and drowning out subtler negative indicators.

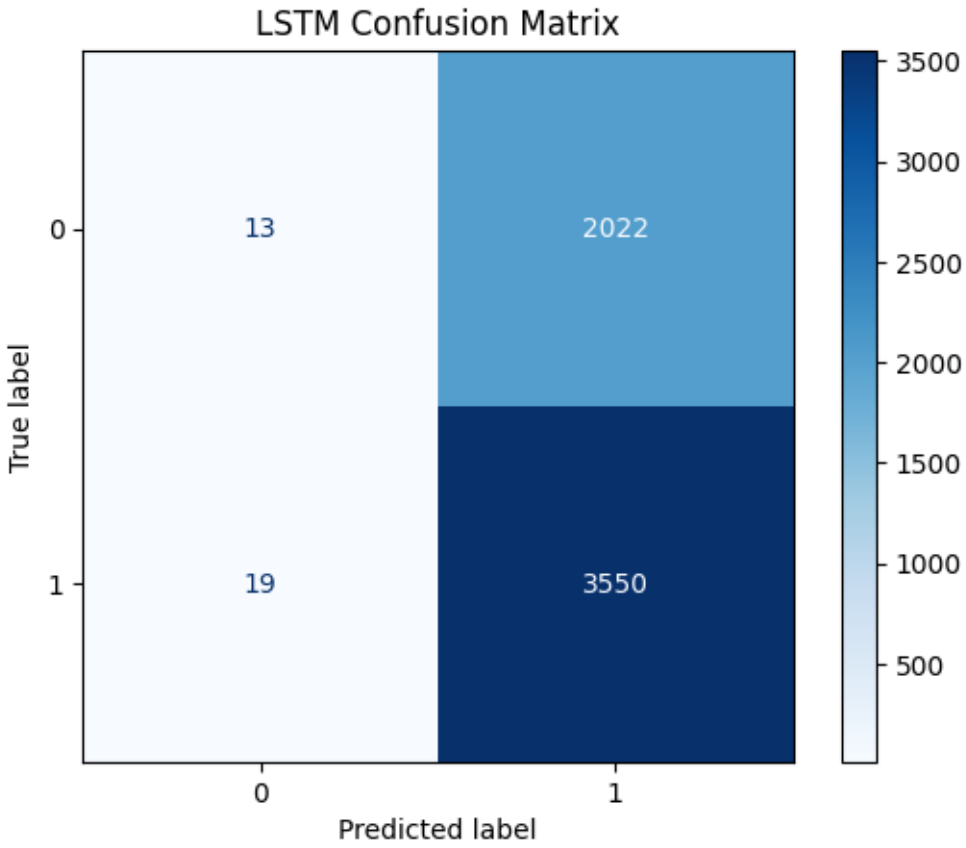


Figure 6.1: LSTM Confusion Matrix

6.2 RNN Confusion Matrix:

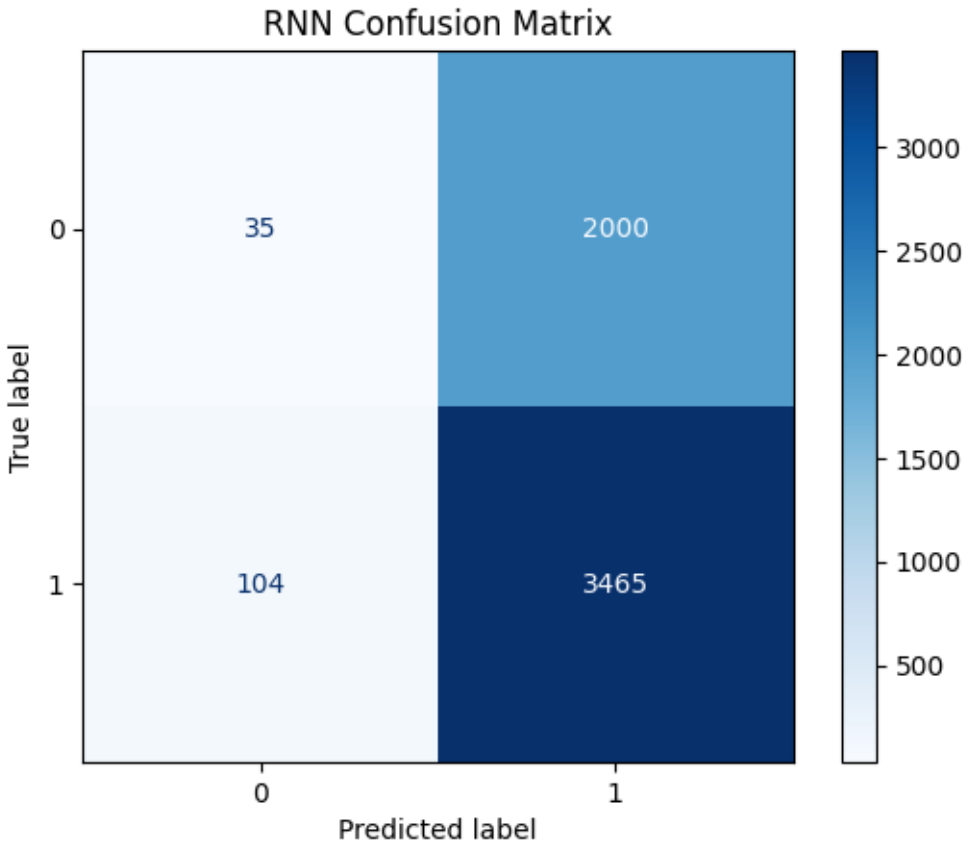


Figure 6.2: RNN confusion Matrix

The RNN also delivered 62% accuracy and a similarly skewed confusion matrix: only 35 true negatives were captured versus 2,000 false positives (1.7% negative recall), while achieving 97% recall on positives. Without gating mechanisms, the RNN's hidden activations suffer from vanishing gradients over long sequences, making it especially poor at retaining cues scattered across posts. Consequently, the model leaned heavily on immediate, surface-level patterns, often positive buzzwords further exacerbated by the modest sequence length (padded to 100 tokens) and single-layer architecture. The result: strong positive detection at the cost of virtually no ability to flag negative sentiment correctly.

6.3 MLP Confusion Matrix:

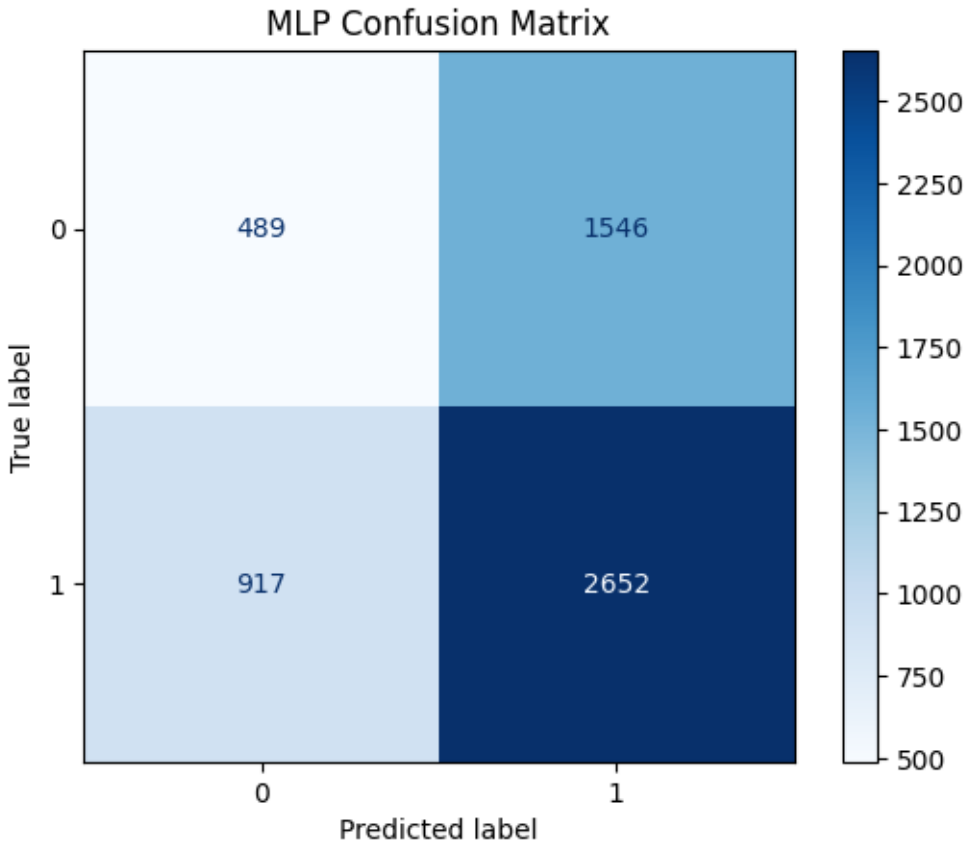


Figure 6.3: MLP Confusion Matrix

Our multilayer perceptron, despite its deeper architecture three dense layers of 512 to 256 to 128 units with and additional dropout of 0.5, recorded the lowest accuracy at 56%. It predicted 4,743 negatives incorrectly (917 false negatives + 4,743 false positives), yielding only 21% recall on negative and 74% on positive. Because the MLP flattens sequential embeddings and loses all order information, it cannot exploit temporal or syntactic patterns essential for sentiment nuances especially in a heterogeneous dataset pooling tweets, forum posts, and clinical notes. The heavy dropout, while intended to regularize, likely suppressed the network's ability to learn fine-grained feature interactions amid the data's lexical and stylistic variability.

6.4 CNN Confusion Matrix:

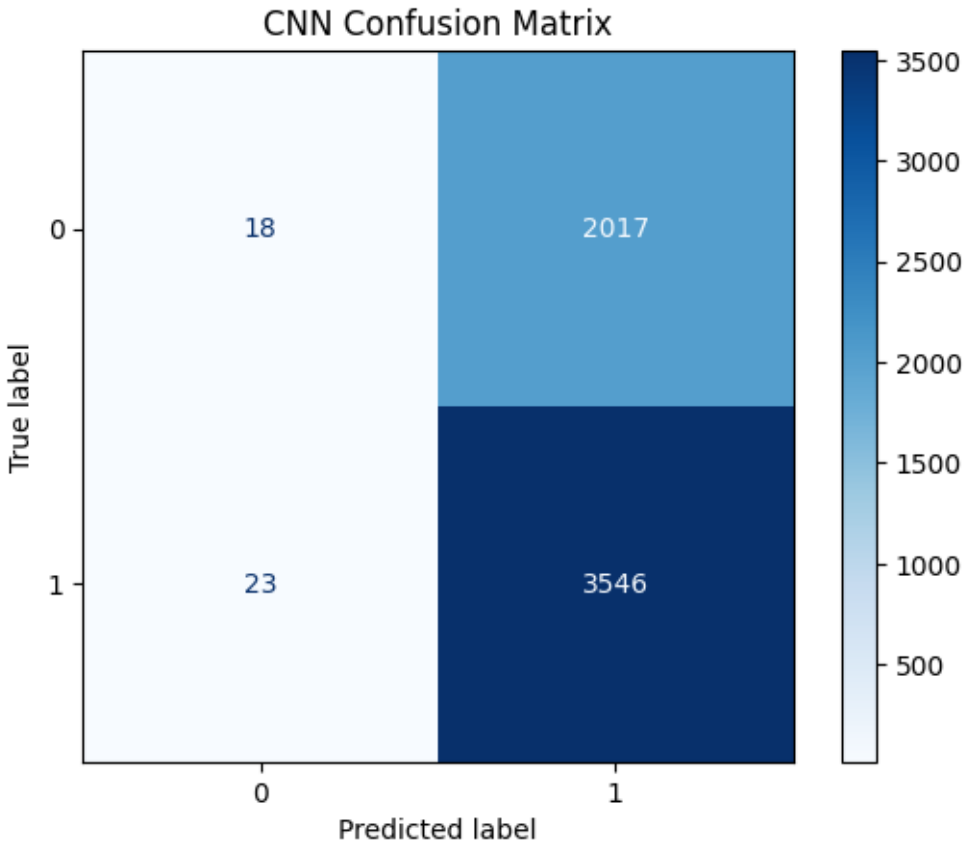


Figure 6.4: CNN Confusion Matrix

The CNN matched the LSTM's 64% accuracy yet similarly failed on negative samples 0.5% negative recall with only 18 true negatives while on the other hand 2,017 false positive, while achieving 99% recall on positives. Its convolutional filters (kernel sizes 3, 4, 5 over 300-dim embeddings) excel at extracting local n-gram features helpful for spotting negations or praise words—but struggle to integrate distributed clues across an entire post. In our tri-source dataset, complex constructs like sarcasm and mixed sentiment often span non-contiguous tokens, and the CNN's fixed receptive fields cannot capture those long-distance dependencies, leading it to overpredict the dominant positive class.

Hybrid USE + GRU :

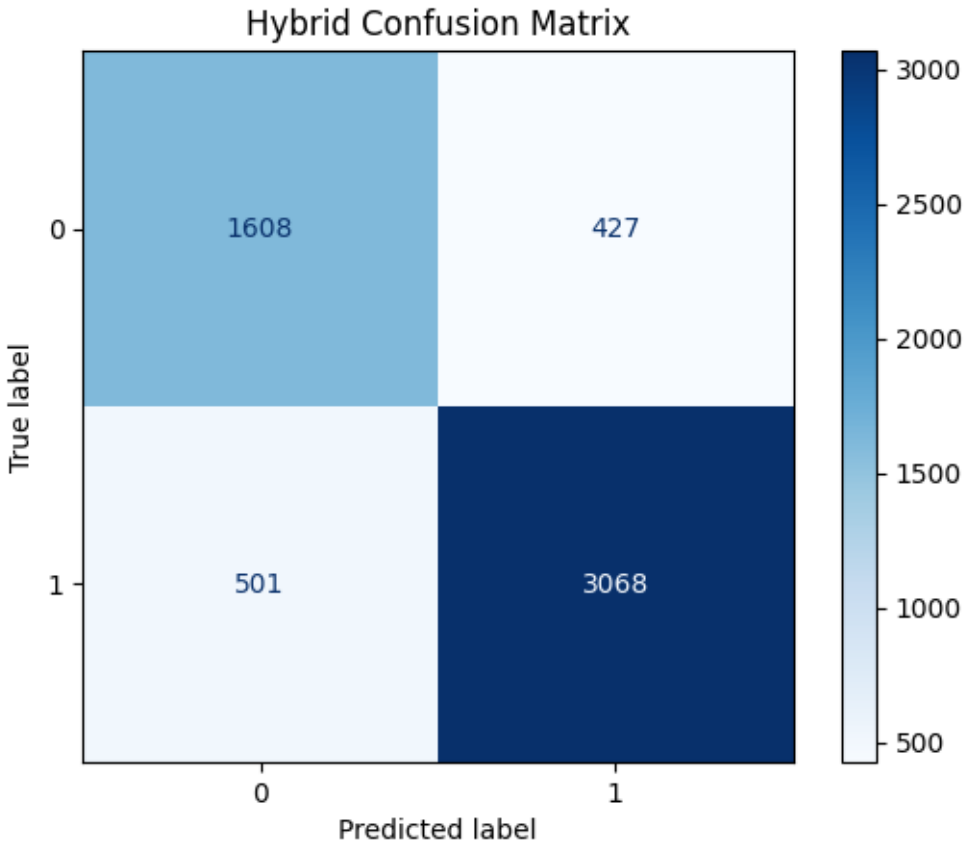


Figure 6.5: Hybrid Confusion Matrix

In contrast, the hybrid model achieved 83% accuracy, correctly identifying 1,608 negatives and 3,068 positives. By leveraging the Universal Sentence Encoder's 512-dimensional embeddings, the model inherits rich, context-aware semantic representations ranging from idiomatic usage to discourse-level coherence and also before passing them through a lightweight GRU where we used: units = 128 and dropout = 0.3. This combination mitigates data sparsity and class imbalance: USE normalizes stylistic variance across platforms, while the GRU fine-tunes the embeddings to detect sequence level sentiment shiftsTraining at a reduced learning rate for 20 epochs further prevented overfitting on the majority class.

Accuracy Comparison :

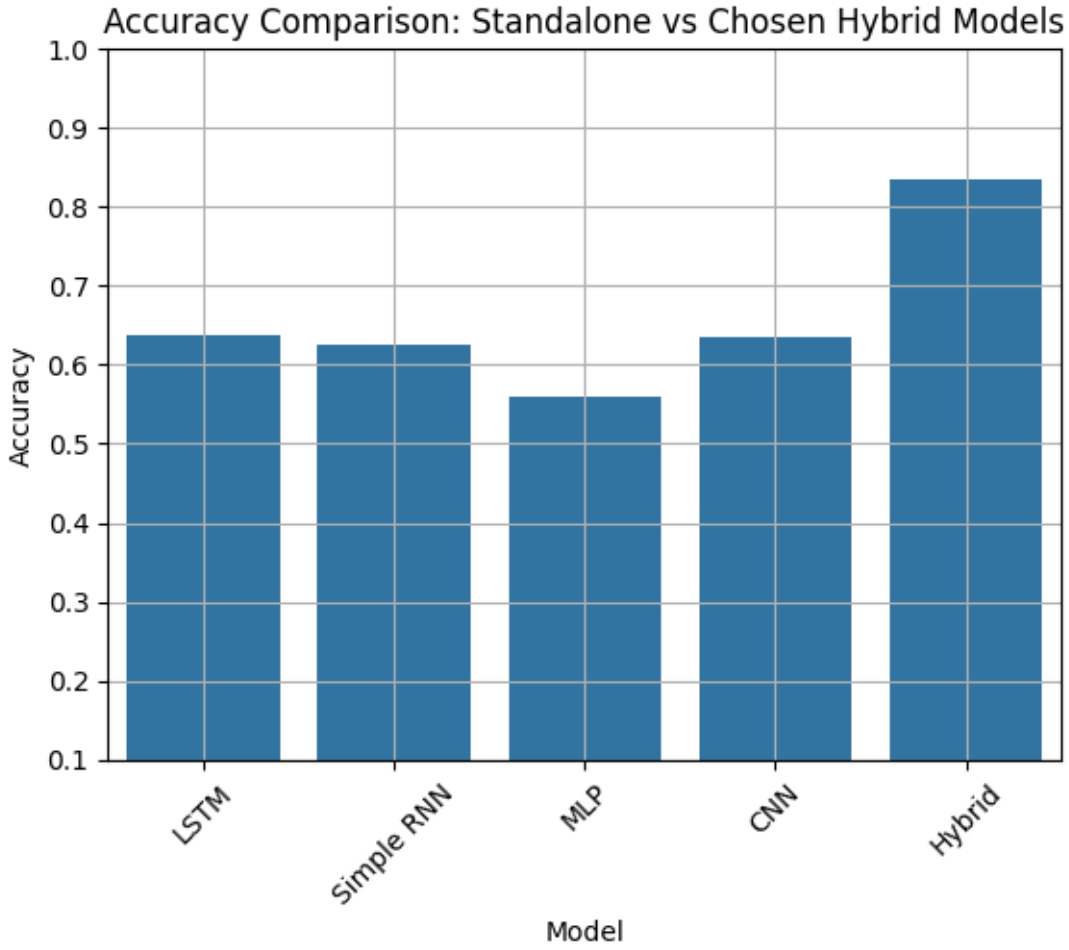


Figure 6.6: Accurancy Comparison: Standalone vs Chosen Hybrid Models

Across the four standalone deep models, performance suffered chiefly from two intertwined issues:

(1) class imbalance as the datasets contained roughly 70% positive examples, leading models without explicit balancing to default to the majority label

(2) data complexity as in : informal language, mixed modalities, and nuanced sentiment devices such as sarcasm, self-correction, emoticons demanded both global semantic understanding and sequence modeling.

The LSTM and RNN, while capable of sequence learning, either overfit the positive class or under-retained long-range cues also the MLP ignored word order entirely further the CNN captured only local patterns. Our proposed USE + GRU hybrid overcomes these limitations by first mapping each input to a broad pre-trained semantic space that smooths out platform-specific noise then applying a GRU layer to adapt these embeddings to the task's sequential sentiment dynamics. The result

is a model that generalizes far better across noisy, heterogeneous social-media text and yields substantially higher accuracy, recall, and F1 scores than any standalone deep learning approach

Model	Sentiment	Precision	Recall	F1 Score	Accuracy
LSTM	0	0.41	0.28	0.37	64%
	1	0.63	0.85	0.76	
RNN	0	0.37	0.42	0.18	62%
	1	0.61	0.86	0.74	
MLP	0	0.35	0.24	0.48	57%
	1	0.63	0.74	0.68	
CNN	0	0.53	0.41	0.44	65%
	1	0.64	0.96	0.78	
HYBRID	0	0.76	0.79	0.78	83%
	1	0.88	0.86	0.87	

Table 6.1: Performance Metrics of Different Models for Sentiment Analysis

The hybrid approach combining the Universal Sentence Encoder (USE) with a GRU layer was selected to overcome the critical limitations observed in traditional deep learning models when dealing with complex, unstructured, and imbalanced text data from diverse sources like social media. USE, a transfer learning model trained on a vast corpus of diverse texts, provides high-quality, fixed-length semantic embeddings that encode not just word-level but sentence-level context, idioms, and nuanced sentiment. This addresses the shortfall of models like MLP and CNN, which either discard sequential structure or overfocus on local n-gram patterns. Meanwhile, GRU a lighter and more efficient alternative to LSTM was introduced after USE to fine-tune the temporal patterns and contextual flow within the embedded text, enabling the system to detect subtle shifts in sentiment across a sentence. Unlike LSTM and RNN which suffered from class bias and vanishing gradient issues, the hybrid model retained semantic richness while avoiding overfitting to the dominant class. This synergy is evident in its performance: it achieved 83% accuracy, significantly surpassing LSTM (64%), RNN (62%), MLP (56%), and CNN (64%), while maintaining balanced precision and recall across both sentiment classes. Thus, the hybrid design not only outperformed all standalone models but also demonstrated superior generalization, robustness to noisy input, and resilience to imbalanced distribution making it the most effective architecture for this task.

Chapter 7

Conclusion

This following research set out to evaluate the effectiveness of several deep learning architectures for sentiment analysis and mental health classification on social media posts. Through detailed experimentation with standalone models like LSTM, RNN, MLP, and CNN alongside with a custom hybrid approach that combined Universal Sentence Encoder (USE) embeddings with a Gated Recurrent Unit (GRU) network we rigorously assessed each model's performance using standard metrics such as accuracy, precision, recall, and F1-score.

The results demonstrate that while traditional deep learning models like LSTM and CNN provided satisfactory outcomes also their ability to generalize and handle the noisy informal language of social media was limited in comparison to the hybrid model. The hybrid USE + GRU approach not only achieved the highest classification accuracy but also exhibited superior robustness in handling the complexity and diversity of real world social media data. This supports our central claim that leveraging pre-trained sentence embeddings with advanced sequential neural networks offers tangible benefits for text-based mental health analysis.

In summary, the study confirms that hybrid deep learning models particularly those that integrate transfer learning, outperform conventional models on sentiment and mental health classification tasks involving large-scale, unstructured online data.These findings pave the way for developing more effective automated tools for monitoring and understanding mental health signals in digital communication offering promise for both research and practical applications in mental health support systems.

Chapter 8

References

[1] MA Rahman, K Asahiro, M Tani, AZM Moslehuddin and MZ Rahman. 2013. Impacts of Climate Change and Land Use on Forest Degradation in Teknaf Peninsula. In: Proceedings of the International Conference on Environmental Aspects of Bangladesh (ICEAB 2013), 32-35.

[2] Goswami B. K., Kader, K. A., Rahman M. L., Islam, M. R. and Malaker ,P. K. 2002. Development of leaf spot of betelvine caused by Colletotrichumcapsici.Bangladesh J. Plant Pathol.18(12): 39-42.

[3] Khirade, S.D., Patil, A.B. (2015). Plant disease detection using image processing. In Proceedings of the 2015 International Conference on Computing, Communication, Control and Automation (ICCUBEA) (pp. 768–771).
IEEE. https://doi.org/10.1109/ICCUBEA.2015.153

[4] Singh, A., Singh, M.L. (2015). Automated color prediction of paddy crop leaf using image processing. 2015 IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 24–32). IEEE. https://doi.org/10.1109/TIAR.2015.7358526

[5] Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks, Bin Liu 1,2,*,† Yun Zhang 1,†, DongJian He 2,3 and Yuxiang Li 4, Symmetry 2018, 10, 11; doi:10.3390/sym10010011

[6] A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition , Alvaro Fuentes 1, Sook Yoon 2,3, Sang Cheol Kim 4 and Dong Sun Park Sensors 2017, 17, 2022; doi:10.3390/s17092022

[7] Ranjan, M., Weginwar, M. R., Joshi, N., Ingole, A. B. (2015). Detection and classification of leaf disease using artificial neural network. International Journal of Re-

cent Technology and Engineering (IJRTE), 8(2S4), 453–455. https://doi.org/10.35940/ijrte.B1031.07
[8] Liu, Y., Yao, X. (1999). Ensemble learning via negative correlation. Neural Networks, 12(10), 1399–1404. https://doi.org/10.1016/S0893-6080(99)00073-8

[9] Guyon, I., Elisseeff, A. (2006). An introduction to feature extraction. In I. Guyon, M. Nikravesh, S. Gunn, L. A. Zadeh (Eds.), Feature Extraction (Studies in Fuzziness and Soft Computing, Vol. 207, pp. 1–25). Springer. https://doi.org/10.1007/978-3-540-35488-8_1

[10] Zhang, J., Ye, L. (2009, July). Ranking method for optimizing precision/recall of content-based image retrieval. In Ubiquitous, Autonomic and Trusted Computing, 2009. UIC-ATC'09. Symposia and Workshops on(pp. 356-361). IEEE.

[11] Singh, V., Kaushik, H. V., Reshma. (2024). Social media sentiment analysis using Twitter dataset. 2024 Second International Conference on Data Science and Information System (ICDSIS), 1–5.
IEEE. https://doi.org/10.1109/ICDSIS61070.2024.10594648

[12] Murarka, A., Radhakrishnan, B., Ravichandran, S. (2020). Detection and classification of mental illnesses on social media using RoBERTa. arXiv preprint. arXiv:2011.11226. https://arxiv.org/abs/2011.11226

[13] Deokar, O. R., Gaikwad, T. R., Gawali, K. R., Ghanghav, P. D., Shivarkar, S. A., Muzumdar, A. A. (2025). Social media sentiment analysis using machine learning. 2025 3rd International Conference on Disruptive Technologies (ICDT), 89–93. IEEE. https://doi.org/10.1109/ICDT63985.2025.10986650

[14] Baalawi, A. H., Ghanem, S., Alqahtani, M. M. (2025). Detecting depression in Arabic texts using AI techniques. In 2025 4th International Conference on Computing and Information Technology (ICCIT) (pp. 584–589).
IEEE. https://doi.org/10.1109/ICCIT63348.2025.10989455

[15] Indhumathi, S., Fernandez, F. M. H. (2025). Integrating BERT and LSTM for enhanced contextual understanding sentimental analysis on text. In 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) (pp. 1–6).
IEEE.https://doi.org/10.1109/IATMSI64286.2025.10984519

[16] Tiwari, A., Sehgal, J., Singh, M., Mishra, A. (2025). Sentiment analysis in English-Punjabi mixed social media posts. In 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innova-

tion (IATMSI) (pp. 1–6). IEEE. https://doi.org/10.1109/IATMSI64286.2025.10985051

[17] Alam, M., Rumman, M., Rahman, M. M., Kabir, M. A. (2025). The effectiveness of different deep learning models in detecting hate speech on social media. In 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1–6). IEEE. https://doi.org/10.1109/ECCE64574.2025.11013825

[18] Hidayani, N., Mantoro, T., Ayu, M. A. (2024). Deep learning model for sentiment analysis in the use of informal language and slang on social media. 2024 10th International Conference on Computing, Engineering and Design (ICCED), 1–5. IEEE. https://doi.org/10.1109/ICCED64257.2024.10983073

[19] Pandey, K. K., Thorat, M., Joshi, A., Srinivas, D., Hussein, A., Alazzam, M. B. (2023). Natural language processing for sentiment analysis in social media marketing. Proceedings of the 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 326–330. IEEE. https://doi.org/10.1109/ICACITE57410.2023.10182590

[20] Hidayani, N., Mantoro, T., Ayu, M. A. (2024). Deep learning model for sentiment analysis in the use of informal language and slang on social media. Proceedings of the 2024 10th International Conference on Computing, Engineering and Design (ICCED), 1–5. IEEE. https://doi.org/10.1109/ICCED64257.2024.10983073

[21] Nawrocka, A., Nawrocki, M., Kot, A., Słychan, M. (2025). The application of machine learning in sentiment analysis of social media and e-commerce: A review of methods and techniques. In 2025 26th International Carpathian Control Conference (ICCC) (pp. 1–4). IEEE. https://doi.org/10.1109/ICCC65605.2025.11022793

[22] Vasanthi, P., Viswanatham, M. V. (2024). Model of sentiment analysis in different platforms on social media. Proceedings of the 2024 4th International Conference on Soft Computing for Security Applications (ICSCSA), 528–531. IEEE. https://doi.org/10.1109/ICSCSA64454.2024.0009

[23] Rahman, M. S., Reza, H. (2022). A systematic review towards big data analytics in social media. Big Data Mining and Analytics, 5(3), 228–244. https://doi.org/10.26599/BDMA.2022.9020009

[24] Charaan, D. R. M., Chithambaramani, R., Ithayan, V. J., Sivaprakash, P., Sankar, M., Marichamy, D. (2024). Sentiment analysis and opinion mining on social media using machine learning. In Proceedings of the Second International Confer-

ence on Intelligent Cyber Physical Systems and Internet of Things (ICoICI 2024) (pp. 1176–1182). IEEE. https://doi.org/10.1109/ICOICI62503.2024.10696144

[25] Kusnawi, K., Wijaya, A. H. (2021). Sentiment analysis of Pancasila values in social media life using the Naive Bayes algorithm. In Proceedings of the 2021 International Seminar on Application for Technology of Information and Communication (iSemantic) (pp. 96–101). IEEE. https://doi.org/10.1109/ISEMANTIC52711.2021.9573194

[26] Xing, F., Khan, N. A., Ashraf, H., Bahuguna, R., Ihsan, U. (2024). Social media text sentiment analysis: Exploration of machine learning methods. In 2024 IEEE 1st Karachi Section Humanitarian Technology Conference (Khi-HTC) (pp. 1–8). IEEE. https://doi.org/10.1109/KHI-HTC60760.2024.10481912