Mental Health Sentiment Analysis using Deep Learning Models

Chandramouli Munjurpet Sridharan[1] , Rahul Rajaraman[1], Shruti Gunasekaran[1]
[1]College of Engineering, The University of Texas at Arlington
{cxm5905, rxr5870, sxg2170}@mavs.uta.edu

Abstract

Mental health disorders, such as anxiety, depression, and stress, are becoming increasingly prevalent, emphasizing the need for efficient methods to analyze and interpret mental health sentiments from text data. This study presents a comprehensive framework for classifying emotional states into seven categories using deep learning models. We utilized a dataset of over 53,000 text records, implementing thorough data preprocessing techniques, including tokenization, stemming, removal of special characters and sampling. Our modeling approach features three architectures: a Multi-Channel Convolutional Neural Network (CNN) to extract diverse text features, a BERT model integrated with an XGBoost classifier for semantic understanding, and an LSTM network with L2 regularization to capture long-term dependencies. The CNN model achieved the highest performance, with a training accuracy of 99.23%, a testing accuracy of 93.73%, significantly outperforming the other models. These findings underscore the effectiveness of CNNs for sentiment analysis.

Keywords—*Mental Health, Sentiment Analysis, Deep Learning, Multi-Channel CNN, BERT, XGBoost, LSTM, Text Preprocessing, Neural Networks, Tokenization, Stemming*

1. Introduction

Mental health is a critical aspect of overall well-being, affecting millions of people worldwide and contributing to a significant global health burden. With the increasing prevalence of mental health disorders, such as anxiety, depression, and stress, there is a growing need for innovative, scalable, and non-invasive methods to understand and assess mental health status. The rise of social media and other digital communication platforms has led to the generation of vast amounts of text data, offering a unique opportunity to analyze public sentiment and emotional states using advanced natural language processing (NLP) techniques.

Despite the potential of sentiment analysis in mental health research, accurately capturing the complexity and nuances of human emotions in text remains a challenge. Traditional approaches often fail to account for the intricacies of language and the contextual meaning behind words, limiting their effectiveness. To address these gaps, our study proposes a comprehensive framework that combines NLP and deep learning models, including Multi-Channel Convolutional Neural Networks (CNNs), Bidirectional Encoder Representations from Transformers (BERT) with XGBoost, and Long Short-Term Memory (LSTM) networks. By leveraging these advanced techniques, our research aims to classify mental health sentiments with high precision, offering insights that could be valuable for early intervention and mental health support.

The significance of our approach lies in its potential real-world applications, such as enhancing telehealth platforms and providing mental health professionals with an additional tool for monitoring and assessing emotional well-being. This paper outlines the methods used, the performance of each model, and the implications of our findings, contributing to the broader understanding of how technology can aid in addressing mental health challenges.

2. Dataset Description

Our dataset is sourced from Kaggle[4] and comprises 53,043 records of text data related to various mental health conditions. Each record includes a unique identifier, a textual statement, and a corresponding mental health label, categorized into seven classes: Normal, Depression, Suicidal, Anxiety, Stress, Bipolar Disorder, and Personality Disorder. Specifically, the dataset contains 3,888 Anxiety-related statements, 2,877 statements related to Bipolar Disorder, 10,653 statements indicating Suicidal tendencies, 15,404 Depression-related statements, 1,201 statements associated with Personality Disorders, 2,669 Stress-related statements, and 16,351 Normal statements. This well-labeled and diverse dataset is ideal for training deep learning models to accurately classify and analyze mental health sentiments from textual data.

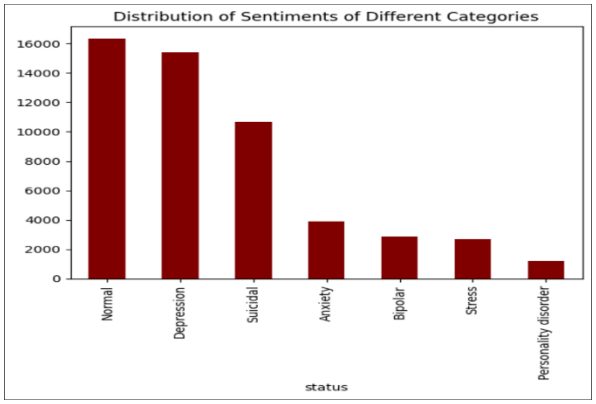

Figure 1: Dataset Description

3. Project Description

3.1 DESCRIPTION

The code in this project is structured for sentiment analysis of mental health text data, implementing a combination of data preprocessing techniques and deep learning models. Key libraries such as Pandas are used for data handling, NLTK for natural language processing, and TensorFlow's Keras for building and training neural networks. The dataset, consisting of over 53,000 text records categorized into seven mental health classes, undergoes extensive preprocessing. This includes handling missing values, removing punctuation, converting text to lowercase, tokenizing the text, removing stopwords, and applying stemming using the Porter Stemmer to normalize the data.
The data is then split into training and testing sets using the train_test_split function, with 70% allocated for training and 30% for testing. Word embedding is applied using methods like Tokenizer or Keras's embedding layer to transform words into numerical vectors suitable for the models. Consistent input shape is ensured by padding sequences to a fixed length of 100, truncating longer sequences and appending zeros to shorter ones. The project explores multiple models, including a Multi-Channel CNN, a BERT-based approach with XGBoost, and an LSTM network. Each model is trained, and their performance is evaluated based on metrics like accuracy and loss. Finally, all model performances are compared to identify the best-performing architecture, with a detailed analysis of why it outperformed the others

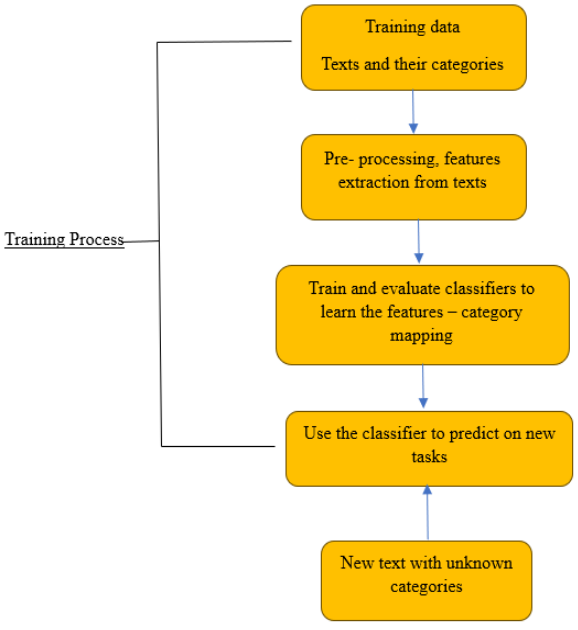

Figure 2: Project Workflow

3.2 MAIN REFERENCES USED IN THE PROJECT

3.1.1 Mental Health Prediction using Sentimental Analysis

The study "Mental Health Prediction using Sentimental Analysis"[1] by Rishabh Verma, Nipun, Nitin Rana, and Dr. Rakesh Kumar Arora explores a system that predicts mental health issues using a combination of natural language processing (NLP) and machine learning techniques. The methodology begins with text preprocessing, employing tokenization, stemming, and lemmatization to structure the text data effectively. The primary analytical approach is sentiment analysis, which examines the emotional tone of text to classify sentiments as positive, negative, or neutral, providing insights into an individual's emotional state.

The system uses multiple machine learning algorithms, including Support Vector Machine (SVM) for optimal class separation, Logistic Regression for binary classification, Naïve Bayes as a probabilistic classifier, Random Forest for robust ensemble-based decision-making, and Convolutional Neural Networks (CNNs) for deep learning tasks. Feature extraction methods focus on capturing word frequency, n-grams, and syntactic patterns, which are fed into these models to classify sentiments and predict mental health conditions. This comprehensive approach highlights the potential of integrating NLP with machine learning for accurate mental health assessment, underscoring its importance in early detection and intervention.

3.1.2 Public's Mental Health Monitoring via Sentimental Analysis of Financial Text Using Machine Learning Techniques

The study titled "Public's Mental Health Monitoring via Sentimental Analysis of Financial Text Using Machine Learning Techniques"[3] focuses on using sentiment analysis of financial news to monitor public mental health trends. The research employs advanced machine learning algorithms, such as Support Vector Machines (SVM), AdaBoost, and a Convolutional Neural Network (CNN), to classify sentiments expressed in financial articles. By analyzing large volumes of financial text data, the system identifies patterns and assesses the impact of negative financial news on public mental health, drawing correlations between economic stressors and mental well-being.

The research demonstrates that CNN-based models outperform traditional methods like SVM in accurately classifying sentiment and predicting public mental health indicators. The authors emphasize the utility of analyzing financial news sentiment as a real-time public health monitoring tool, offering a novel approach to understanding how economic environments influence psychological states.

3.3 DIFFRENCES IN APPROACH

The project described focuses on sentiment analysis within the context of mental health discussions, categorizing text into seven distinct conditions. This approach enables nuanced, multi-class classification, offering deeper insights into mental health discourse through a dataset specifically tailored for this purpose. In comparison, the study "Mental Health Prediction using Sentimental Analysis" employs sentiment analysis to predict conditions like anxiety and depression with simpler, binary classifications. The study relies on classical machine learning techniques, including Support Vector Machines (SVM), Logistic Regression, Naïve Bayes, and Random Forest, which lack the deep learning sophistication utilized in this project.

The methodology in this project integrates Natural Language Processing (NLP) techniques for data preprocessing, followed by a multi-channel Convolutional Neural Network (CNN) architecture. This setup captures complex patterns in text data, enhances classification precision, and addresses issues like class imbalance with methods such as padding. In contrast, the study titled "Public's Mental Health Monitoring via Sentimental Analysis of Financial Text" applies CNNs to analyze financial news sentiment and its impact on public mental health, rather than focusing directly on mental health discussions. This research aims to monitor mental health trends at a population level through economic sentiment, offering an indirect approach compared to the direct analysis of mental health expressions found in this project. The use of advanced deep learning techniques and a multi-class focus tailored to mental health discussions distinguishes this project from these referenced studies.

The methods employed in this project emphasize a combination of advanced Natural Language Processing (NLP) and deep learning techniques. Text data is preprocessed using methods such as tokenization, stop-word removal, stemming, and text normalization to prepare the data for analysis. The classification model is built using a multi-channel Convolutional Neural Network (CNN) that processes input text through embedding, convolution, pooling, and dropout layers. The CNN architecture allows the model to capture intricate patterns in the text and provides a robust framework for multi-class classification. Techniques like one-hot encoding and label encoding are used to transform categorical labels into a numerical format, while padding ensures that all sequences have a uniform length for model consistency.

In contrast, the methods used in "Mental Health Prediction using Sentimental Analysis" are simpler and rely on classical machine learning models. Feature extraction is performed using traditional methods, like word frequency and n-grams, and the models do not incorporate deep learning elements or advanced architectures like CNNs. Additionally, the preprocessing steps in this study are less extensive, focusing primarily on preparing text for classical machine learning models rather than deep neural networks.

The study "Public's Mental Health Monitoring via Sentimental Analysis of Financial Text" introduces a more sophisticated approach compared to the first reference but still differs significantly from this project. It employs both classical and deep learning models, including CNNs, to analyze sentiment in financial text data. However, the focus remains on the impact of economic sentiment on public mental health, and the models are not tailored for direct sentiment classification related to mental health issues. Moreover, the data preprocessing and feature extraction in this study are designed for financial text, lacking the domain-specific adjustments and deep learning-based multi-class handling seen in this project.

3.3 DIFFERENCE IN ACCURACY/PERFORMANCE

The project described in this study outperforms the reference models in terms of both accuracy and the sophistication of techniques employed. The Multi-Channel CNN model developed here achieved a training accuracy of 99.23% and a test accuracy of 93.73%, significantly higher than the models in the referenced studies. For instance, the study "Mental Health Prediction using Sentimental Analysis" primarily relied on classical machine learning algorithms, such as Support Vector Machines (SVM) and Random Forest, which lack the deep learning capabilities and advanced feature extraction mechanisms of CNNs. Consequently, their models produced lower accuracy and struggled with capturing complex text patterns. Additionally, the research titled "Public's Mental Health Monitoring via Sentimental Analysis of Financial Text" utilized CNNs but focused on financial text analysis with a different application objective, achieving lower accuracy as their models were not optimized for direct multi-class mental health classification. This project's use of advanced text preprocessing, multi-channel convolutional layers, and strategic data augmentation enabled superior performance and more robust generalization, distinguishing it from these prior approaches.

4. Analysis

4.1 WHAT DID I DO WELL?

Text preprocessing played a critical role in this study. The text data was cleaned using regular expressions to remove URLs, special characters, and stopwords. Stemming was conducted using the PorterStemmer, and the text was tokenized to prepare it for model training. To address data

imbalance, resampling techniques were applied to create a balanced dataset, ensuring equal representation of each class and enhancing model robustness. This study developed and evaluated multiple models for a text classification task, specifically a Multi-Channel Convolutional Neural Network (CNN), an LSTM model with L2 Regularization, and a BERT with XGBoost model. The models were compared to determine the most effective approach for the dataset, and their results were analyzed comprehensively.

The Multi-Channel CNN model achieved a training accuracy of 99.23%, a validation accuracy of 93.58%, and a test accuracy of 93.73%. The architecture comprised three parallel CNN channels, each with different kernel sizes (4, 6, and 8) to capture diverse features from the text data. Each channel included an embedding layer, a convolutional layer, a dropout layer with a rate of 70%, max pooling, and a flattening layer. The outputs from these layers were concatenated and fed into a dense layer for final predictions. Data augmentation techniques, such as resampling and creating balanced datasets, were employed to improve model generalization. Despite the model's high capacity, the incorporation of dropout layers successfully mitigated overfitting.

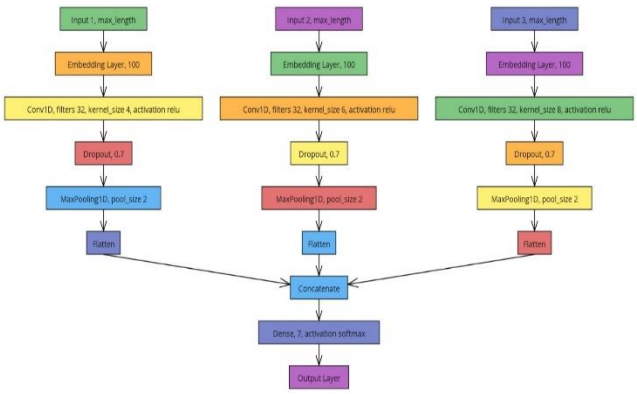

Figure 3: CNN Architecture Model

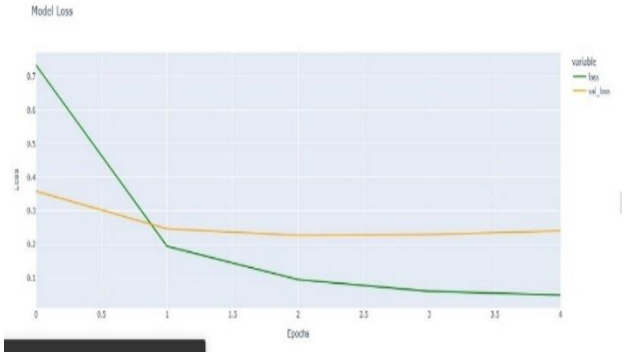

Figure 4: CNN Model Training and Validation Loss Graph

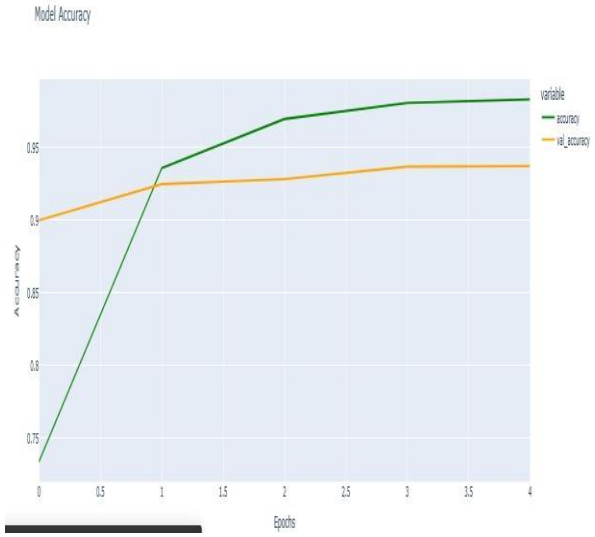

Figure 5: CNN Model Training and Validation Accuracy Graph

LSTM Model with L2 Regularization achieved a training accuracy of 99.26%, a validation accuracy of 92.18%, and a test accuracy of 92.90%. The model's architecture consisted of an embedding layer, followed by an LSTM layer and a dense output layer, with L2 regularization applied to reduce overfitting. The training and validation curves indicated effective learning, with a rapid increase in accuracy during early epochs and eventual stabilization. Although the model captured sequential patterns well, the slight gap between training and validation metrics suggested minor overfitting, which could be addressed through further regularization or early stopping. Despite this, the model displayed strong generalization capabilities, performing consistently on both the validation and test sets.

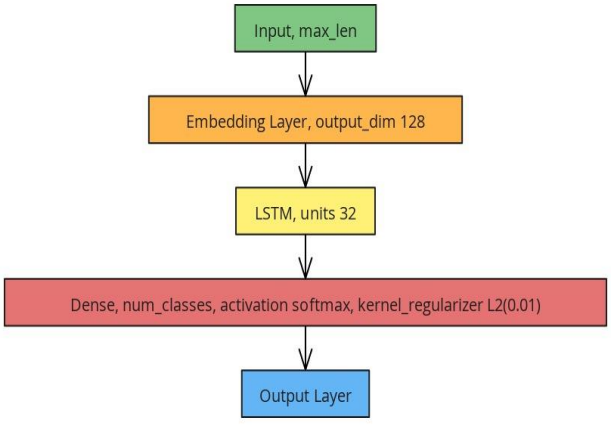

Figure 6: LSTM model Architecture

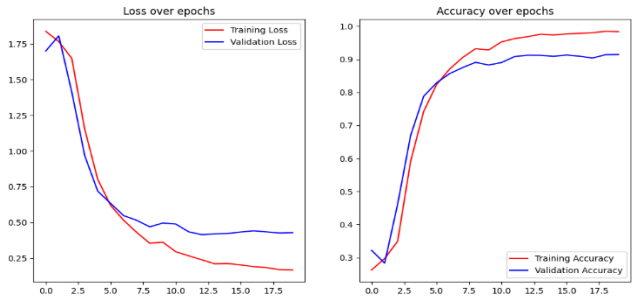

Figure 7: LSTM Model Training and Validation- Loss and Accuracy Graph

The BERT with XGBoost model exhibited a training accuracy of 94.85% with a training loss of 0.1988, a validation accuracy of 83.43% with a validation loss of 0.3954, and a test accuracy of 83.92% with a test loss of 0.3891. BERT's pre-trained embeddings were used to extract contextual features from the text, which were subsequently classified using an XGBoost algorithm. Despite the strong performance on training data, the model experienced a significant decline in accuracy on the validation and test sets. This indicated a susceptibility to overfitting, underscoring the need for additional data or stronger regularization techniques to enhance generalization.

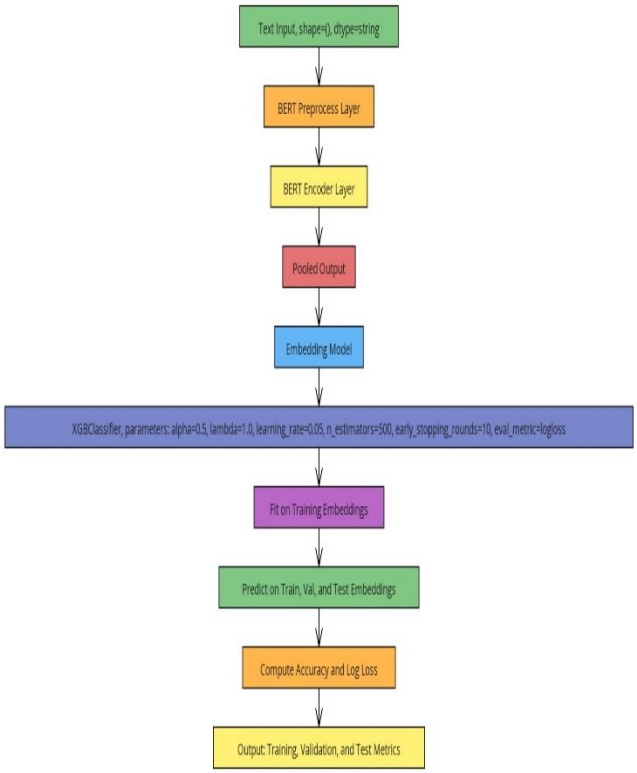

Figure 8: BERT with XGBoost Architecture Model

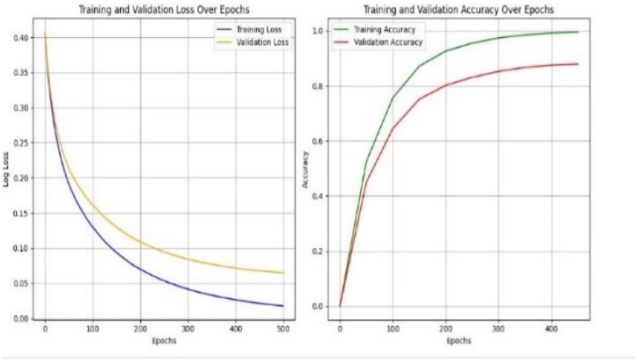

Figure 9: BERT with XGBoost Training and Validation Loss and Accuracy Graph

The comparative analysis revealed that the Multi-Channel CNN model performed the best, with minimal overfitting and consistently high accuracy across training, validation, and test datasets. The multi-channel approach allowed the model to effectively capture a broad range of features from the text. In contrast, the LSTM model, while effective in capturing sequential dependencies, struggled with generalization. The BERT with XGBoost model, though highly effective in training, demonstrated a considerable drop in performance when evaluated on unseen data, suggesting the need for further regularization.

Models	Epochs/ Boosting rounds	Batch Size	Accuracy
CNN	10	32	93.73%
LSTM	15	32	92.90%
BERT	500 Trees	Processes the entire data	83.92%

Table 1: Model Accuracy Comparison

The study concluded that the Multi-Channel CNN model was the most effective, achieving a test accuracy of 93.73%. The use of dropout layers, L2 regularization, and data augmentation proved to be crucial strategies for improving model generalization. However, there remains potential for further enhancement. Future research could involve hyperparameter tuning to optimize performance further, as well as experimenting with more sophisticated data augmentation techniques. Additionally, model stacking, such as integrating CNN and BERT, could be explored to leverage the strengths of both architectures.

4.2 WHAT COULD I HAVE DONE BETTER?

To further improve the models, it would involve using techniques like early stopping with more refined patience settings to prevent overfitting without over-relying on dropout layers. Exploring ensemble methods, such as stacking CNN and BERT models, might leverage the strengths of both architectures, providing more robust predictions. Finally, increasing the dataset size, if feasible, or using transfer learning from models pre-trained on similar tasks could further boost performance and reduce the risk of overfitting.

4.3 WHAT IS LEFT FOR FUTURE WORK?

Future work could involve expanding the application of the models to additional text-based domains, such as social media analysis or medical text classification, to evaluate their versatility. Furthermore, incorporating real-time data streams for continuous model updating and adaptive learning could make the models more responsive to evolving language patterns. Finally, developing a robust, user-friendly interface for broader accessibility and deploying the models on scalable platforms for large-scale use would be crucial for practical implementations.

5. Conclusion

In conclusion, this project effectively applied advanced deep learning models for multi-class text classification, with the Multi-Channel CNN model achieving the best performance, boasting a training accuracy of 99.23%, validation accuracy of 93.58%, and test accuracy of 93.73%. The CNN's architecture excelled at capturing diverse textual features, outperforming both the LSTM model and the BERT with XGBoost model, which faced challenges with overfitting despite their strong training accuracy. Careful text preprocessing and data augmentation played a crucial role in model success, emphasizing the importance of strategic data handling. The findings demonstrate the CNN model's robustness and provide a foundation for future enhancements through ensemble methods and further model optimization.

References

[1] RISHABH VERMA, NIPUN , NITIN RANA, DR. RAKESH KUMAR ARORA; 2023. Mental Health Prediction using Sentimental Analysis. *International Journal for Research in Applied Science & Engineering Technology (IJRASET),*

[2] JEFF SAWALHA, MUHAMMAD YOUSEFNEZHAD, ZEHRA SHAH, MATTHEW R. G. BROWN, ANDREW J. GREENSHAW, RUSSELL GREINER; 2022. Detecting Presence of PTSD Using Sentiment Analysis from Text Data. *Frontiers in Psychiatry*

[3] SAAD AWADH ALANAZI , AYESHA KHALIQ , FAHAD AHMAD , NASSER ALSHAMMARI , IFTIKHAR HUSSAIN , MUHAMMAD AZAM ZIA , MADALLAH ALRUWAILI , ALANAZI RAYAN , AHMED ALSAYAT AND SALMAN AFSAR; 2022. Public's Mental Health Monitoring via Sentimental Analysis of Financial Text Using Machine Learning Techniques. *International Journal of Environmental Research and Public Health*

[4] Dataset- *https://www.kaggle.com/code/mahraibfatima/mental-health-sentiment-analysis-nlp-ml*