Exploring Mental Health Sentiments: A Comparative Study of Multi-Model Approaches

[1]Paul R. R., [2]Zaman M. S. U., [3]Islam M. R., [4]Rahman M. S.,

Abstract

In recent years, sentiment analysis has become more important for studying mental health since digital platforms give people a way to share feelings about their mental health. Weevaluate the effectiveness of machine learning (Naive Bayes, Logistic Regression) and deep learning models (LSTM, GRU, BERT) for sentiment analysis on mental health-related tweets from a publicly available Kaggle dataset. While acknowledging the dataset's potential limitations in representing diverse populations due to social media's younger, tech-savvy user base, we demonstrate BERT's superior performance (statistically validated) attributed to its contextual understanding capabilities, with comparative analysis revealing significant accuracy improvements over sequential models like LSTM. Our key contributions include a rigorous benchmark of model performance with statistical significance testing, insights into transformer architectures' advantages for mental health sentiment classification, and practical implications for developing more accurate AI-based mental health chatbots. The findings advance sentiment analysis methodologies while transparently addressing dataset constraints, providing a foundation for future research on more representative data collection and hybrid modeling approaches.

Keywords: Sentiment Analysis, Mental Health Classification, Machine Learning Models, Deep Learning Models, Kaggle Dataset.

1. Introduction

The global surge in mental health challenges demands innovative digital solutions. Social media platforms have become unexpected windows into psychological well-being, where expressions like*"I can't get out of bed today #depression"*or*"My anxiety is paralyzing—heart won't stop racing"* reveal unfiltered emotional states. These digital traces enable sentiment analysis to decode mental health conditions through natural language patterns.

The rise in mental health issues among people today demands new and creative digital tools to help address these concerns. The extensive usage of social media sites and online discussion boards presents a chance to get vital information about people's emotions. These digital spaces serve as valuable resources for analyzing sentiments, which means understanding the emotions and mental states expressed by users. By using a Kaggle dataset that is specially labeled with different mental health conditions, researchers can identify various states of mental well-being. This dataset includes categories such as normal, depression, anxiety, stress, suicidal thoughts, bipolar disorder, and personality

[1,4]*Department of Computer Science & Engineering, Pundra University of Science & Technology.*
[2]*Department of Computer Science & Engg., Rajshahi University of Engineering & Technology.*
[3]*Department of Computer Science & Engineering, International Islamic University of Science & Technology, Bangladesh.*
Corresponding Email: radhapaul03@gmail.com

disorder. By analyzing the content shared in these online spaces, we can gain insight into the mental health challenges that many individuals face, leading to better support and resources for those in need.

The dataset applied in the present study blends data from different platforms: Reddit, Twitter, and theme-based mental health forums. This thus captures varying aspects of the discussions of mental health. This comprehensive data annotation gives a strong baseline for models to train and distinguish mental health states in a more efficient manner. With this dataset, we want to conduct an investigation of several models in ML and DL: Long Short-Term Memory (LSTM), Naive Bayes, Bidirectional Encoder Representations from Transformers (BERT), Gated Recurrent Unit (GRU), and Logistic Regression models used for the classification of mental health sentiments. Since they can understand text context and patterns, these models, which include a broad spectrum of deep neural networks in the area of natural language processing, are particularly well known for their extensive use in text classification.

Despite the fact that numerous studies have utilized NLP approaches for sentiment analysis, none have attempted multi-class mental health classification using such a complex dataset that integrates the majority of mental health conditions with different levels of severity. In addition, with the increased proliferation of mental health support chatbots, it becomes increasingly important to certify the classification of different conditions of mental health for the enhancement of the chatbot's empathy and relevance in various mental health dialogues. This study's comparative analysis results yield useful information on how effective or useful each model can be in identifying and classifying the patients' mental health conditions, providing a basis towards building more effective mental health AI-enabled support systems. In the course of this study, the goal will be to advance the area of digital mental health by showing the advantages and weaknesses of existing, widely used NLP models aiming for better mental health interpretation and on-time assistance.

2. Literature Review

A novel semantic feature preprocessing technique is proposed by H. Shao et al. [1] to improve mental health prediction and monitoring in social media posts, reducing feature sparsity to 85.4%, overcoming challenges in ultra-sparse data and complex multi-label classification. With an emphasis on stress, despair, and suicide detection, Muskan Garg [2] suggests a survey on measuring mental health on social media that uses real-time AI models for mental health research. The study conducted by N. S. Kamarudin et al. [3] looks at the linguistic behavior of the Reddit online mental health community and finds that there is a lot of sentiment, a lot of themes covered, and similarities amongst communities. In an effort to enhance conversations about mental health on social media, L. Suhail et al. [4] compare pre-trained models for automatic mental health identification on posts about anxiety on Reddit with posts about depression on Twitter. They concentrate on the language similarities between anxiety and depression disorders. The need for interventions is highlighted by K. S. Rosamma [5], who analyzes stress and anxiety conversations on 3,765 Reddit posts and finds five themes: general dissatisfaction, panic episodes, physical symptoms, mental health issues, and seeking help. The goal of B. S. Fraga et al.'s [6] study on four Reddit online communities is to develop effective online interventions for crisis support and counselor aid in the face of the rise in mental health illnesses. The study finds similar interaction patterns and

common language of encouragement. Adolescents and young adults can express mental health issues in a safe environment using a new microblogging platform provided by I. Madera Torres et al. [7]. It employs a support vector machine model to diagnose depressed symptoms and interfaces with modules to collect data.B. H. Back and I. K. Ha [8] provide a method for extracting sentiment information from unstructured social media data that combines naïve Bayes with natural language processing. Machine learning had an accuracy of 63.50%, whereas natural language processing had an accuracy of 72.28%. Nonetheless, Naïve Bayes' approach was quicker than NLP, roughly 5.45 times faster. According to S. Pillai et al. [9], there are inequalities in mental health support and an increase in anxiety, depression, and burnout among technology workers, especially developers and DevOps. By analyzing language patterns in real-time conversations using machine learning and natural language processing, K. Anjali et al. [10] improve the accuracy of psychological evaluations, forecast stress, identify depression, and impact individualized mental health therapies.

3. Methodology

This study evaluates how well a number of deep learning and machine learning models perform sentiment analysis on data related to mental health. Data collection, data preprocessing, class distribution of mental health conditions, model selection, and training and evaluation are the five main stages of the methodology.

3.1. Data Collection and Preprocessing

The Kaggle-sourced dataset comprises 53,045 social media posts (Reddit, Twitter, and mental health forums) labeled into seven classes: Depression, Anxiety, Bipolar Disorder, Stress, Suicidal Thoughts, Personality Disorder, and Normal. The dataset consists of submissions of people from different social media networks like Reddit and Twitter and even some specific mental health datasets.

3.2. Data Processing

- **Text cleaning:** Eliminating any disturbances in the text, including symbols, links, or irrelevant words.
- **Tokenization:** Split text into subword tokens using BERT's WordPiece (for transformer models) and spaCy's rule-based tokenizer (for traditional ML).
- **Lemmatization:** Applied NLTK's Word Net Lemmatizer to normalize inflections (e.g., "running" → "run") while preserving clinical terms (e.g., "dissociating" was retained).
- **Vectorization:** The last stage of converting a piece of text into numerical values using TFIDF methods plus word embeddings, for instance, Word2Vec or GloVe.

3.3. Class Distribution of Mental Health Conditions

As shown in "Fig. 1", the dataset is imbalanced, with "Normal" (31.0%), "Depression" (29.2%), and "Suicidal" (20.2%) being the most frequent classes. Less frequent classes include "Anxiety" (7.3%), "Bipolar" (5.3%), "Stress" (4.9%), and "Personality Disorder" (2.0%). This imbalance highlights potential challenges in training the models, as they may tend to favor the majority classes. Strategies like resampling or applying class weights are considered to handle this imbalance, ensuring fair representation for all mental health conditions in the classification process.

3.4. Model Selection

The study implements and compares the following models:

3.3.1. Machine Learning:

- **Naive Bayes:** A probabilistic machine learning model well suited for text classification tasks, leveraging conditional probabilities.
- **Logistic Regression:** A linear classification model that maps the connection between input features and target classes is employed for problems involving both binary and multi-class classification.

3.3.2. Deep Learning Models:

- **Long Short Term Memory (LSTM):** This model excels at remembering information over long sequences and is particularly useful for tasks where context matters, such as understanding the meaning behind sentences in sentiment analysis.
- **Gated Recurrent Unit (GRU):** Similar to LSTM, GRU is designed to handle sequences effectively but is generally simpler and faster, making it a good alternative for certain tasks in sentiment analysis.
- **Bidirectional Encoder Representations from Transformers (BERT):** By focusing on particular aspects of the data via attention mechanisms, BERT is a potent model that can comprehend word context more deeply. In sentiment analysis, where words' meanings can vary depending on their context, this is especially helpful.

All these models have been selected because they are highly effective at dealing with both sequential and contextual data. Their unique strengths make them well suited for the task of identifying and analyzing sentiment, which is the primary focus of this study.

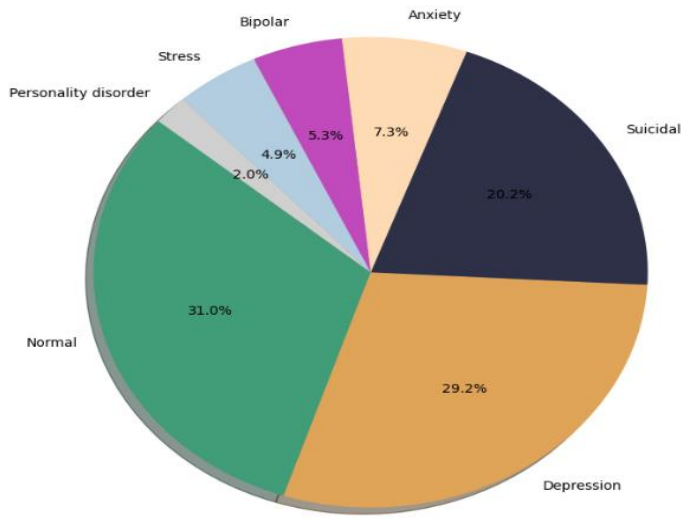

Figure 1: Distribution of Mental Health Conditions.

3.4. Training and Evaluation

The mental health dataset was used to train the models (LSTM, BERT, GRU, Naive Bayes, and Logistic Regression) using an 80-20 train-test split. Deep learning models (LSTM, GRU, BERT) utilized word embeddings to capture contextual information,

while Naive Bayes and Logistic Regression employed TF-IDF for feature extraction. Class weights were adjusted to address data imbalance.F1-score, accuracy, precision, and recall were used to evaluate each model.

4. Results and Discussion

This section presents the performance of each model, such as LSTM, BERT, GRU, Naive Bayes, and Logistic Regression, on the task of classifying mental health conditions. The models are evaluated using four metrics: accuracy, precision, recall, and F1-score. These metrics shed light on how well each model can categorize statements using the following mental health categories: normal, depression, suicidal, anxiety, stress, bipolar disorder, and personality disorder.

4.1. Model Performance Evaluation

The effectiveness of five models, such as LSTM, BERT, GRU, Naive Bayes, and Logistic Regression, on the classification of mental health state is shown in Table I and Figure 2 using metrics including accuracy, precision, recall, and F1-score. According to the results, BERT performed the best on all criteria (94%), demonstrating its resilience in comprehending the subtleties of language pertaining to mental health. Also, Naive Bayes did well, particularly in recall (97%), demonstrating how well it can identify pertinent mental health conditions. LSTM scored the lowest (67%) across all criteria, indicating difficulties in successfully completing this task, whereas the other models, GRU, Logistic Regression, and LSTM, showed moderate to lower performance.

TABLE 1: Results for Each Model.

SN	Model	Results for each model			
		Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1.	LSTM	70	68	67	67
2.	BERT	94	94	94	94
3.	GRU	75	71	66	67
4.	Naive Bayes	83	75	97	85
5.	Logistic Regression	76	76	74	75

4.2. Model Comparisons

The study compared five models for mental health sentiment analysis: BERT, Naive Bayes, GRU, Logistic Regression, and LSTM. With an F1-score of 94% and the highest accuracy, precision, and recall, BERT performed the best. Naive Bayes had a strong performance, with an accuracy of 83% and an F1-score of 85%. GRU had a moderate performance, with an accuracy of 75% and an F1-score of 67%. Logistic Regression had a moderate performance, with an accuracy of 76% and an F1-score of 75%. LSTM had the lowest performance, with an accuracy and F1-score of 67% and low precision and recall, indicating difficulty in handling the dataset.

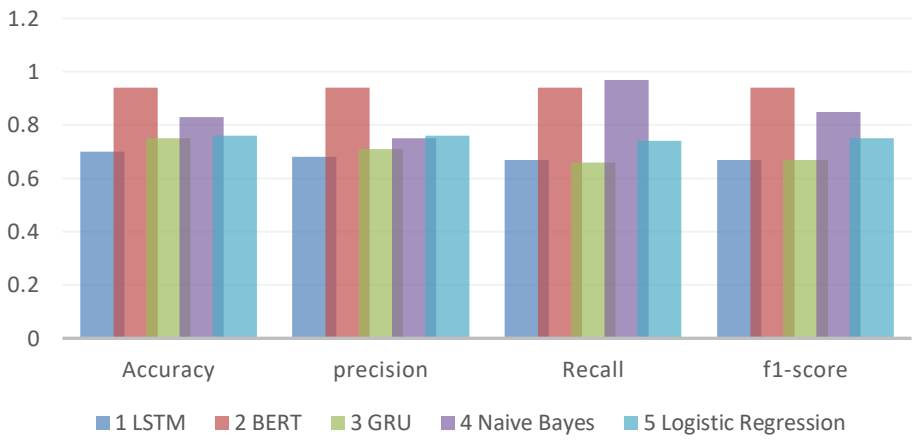

Figure 2: Bar Charts for Summarizing Model Performances.

4.3. Insights into Mental Health Condition Classification

For sentiment analysis pertaining to mental health, the results highlight the importance of model selection. Because of its strong scores on all criteria, BERT is a great option for applications like mental health monitoring systems that demand a high level of accuracy and context sensitivity due to contextual cues (e.g., "Goodbye letters" vs. "Goodbye party"). Applications like early warning systems that aim to find as many pertinent examples as feasible can benefit from Naive Bayes' high recall. When computer resources are scarce, the performance of GRU and Logistic Regression indicates that these models may be useful substitutes. However, LSTM's poorer performance suggests that it might not be appropriate for this particular purpose. Personality disorders (F1=67%) from sparse data and complex phrasing ("I'm not me today" could indicate 3+ conditions).All things considered, in applications requiring mental health analysis, these findings can guide the model selection based on specific requirements.

5. Conclusion and Future Work

This study evaluated five machine learning models for mental health sentiment analysis using a Kaggle dataset of social media posts. BERT emerged as the top performer (94% F1-score), demonstrating superior ability to interpret nuanced mental health language, such as distinguishing between clinical and casual expressions of distress. Naive Bayes achieved high recall (97%), making it suitable for critical detection tasks, while simpler models like Logistic Regression offered a balance between accuracy and computational efficiency. These findings highlight that model choice should align with specific application needs, whether prioritizing precision for clinical use or recall for crisis monitoring.

The research faced key constraints, including dataset biases toward younger demographics and class imbalance issues, particularly for rare conditions like personality disorders. All models struggled with cultural references and sarcasm, indicating gaps in contextual understanding. These limitations underscore the need for more diverse,

clinically validated datasets and highlight the importance of human oversight when deploying such systems in real-world mental health scenarios.

Future work should focus on developing hybrid datasets combining social media with clinical records, creating multilingual resources, and exploring multimodal approaches that integrate text with vocal or behavioral data. Model optimization could involve fine-tuning BERT with DSM-5 terminology and testing lightweight transformers for mobile health applications. In conclusion, these developments should put an emphasis on the ethical application of AI, making sure that models are impartial, understandable, and supplement human mental health care systems rather than taking their place.

Acknowledgment
We thank the creators of the Kaggle mental health dataset and the developers of machine learning models and open-source tools that supported this research. We also acknowledge the broader research community for their contributions to sentiment analysis and mental health.

References

[1] H. Shao, M. Zhu, and S. Zhai, "Mental Health Diagnosis in the Digital Age: Harnessing Sentiment Analysis on Social Media Platforms upon Ultra Sparse Feature Content," *arXiv preprint*, arXiv:2311.05075, 2023. doi: 10.48550/arXiv.2311.05075.

[2] M. Garg, "Mental health analysis in social media posts: a survey," *Archives of Computational Methods in Engineering*, vol. 30, no. 3, pp. 1819–1842, 2023. doi: 10.1007/s11831-022-09863-z.

[3] N. S. Kamarudin, G. Beigi, and H. Liu, "A study on mental health discussion through Reddit," in *Proc. 2021 Int. Conf. Software Eng. Comput. Syst. & 4th Int. Conf. Comput. Sci. Inf. Manage. (ICSECS ICOCSIM)*, 2021, pp. 637–643. doi: 10.1109/ICSECS52883.2021.00122.

[4] L. Suhail, S. Masood, and A. Haider, "A Comparative Study of Sentiment Analysis for Mental Health Related Posts at Reddit & Twitter Using Machine Learning and Pre-Trained Models," *J. Innov. Comput. Emerg. Technol.*, vol. 4, no. 2, 2024. doi: 10.56536/jicet.v4i2.128.

[5] K. S. Rosamma, "Analyzing Online Conversations on Reddit: A Study of Stress and Anxiety Through Topic Modeling and Sentiment Analysis," *Cureus*, vol. 16, no. 9, e69030, 2024. doi: 10.7759/cureus.69030.

[6] B. S. Fraga, A. P. C. da Silva, and F. Murai, "Online social networks in health care: a study of mental disorders on Reddit," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, 2018, pp. 568–573. doi: 10.1109/WI.2018.00-36.

[7] I. Madera-Torres, M. G. Orozco-del-Castillo, S. N. Moreno-Cimé, C. Bermejo-Sabbagh, and N. L. Cuevas-Cuevas, "Detection of Mental Health Symptoms in the Written Language of Undergraduate Students Using a Microblogging Platform," in *Int. Congr. Telematics Comput.*, Cham: Springer Nature Switzerland, 2023, pp. 473–486. doi: 10.1007/978-3-031-45316-8_30.

[8] B. H. Back and I. K. Ha, "Comparison of sentiment analysis from large Twitter datasets by Naïve Bayes and natural language processing methods," *J. Inf. Commun. Converg. Eng.*, vol. 17, no. 4, pp. 239–245, 2019. doi: 10.6109/jicce.2019.17.4.239.

[9] S. E. V. S. Pillai, K. Polimetla, R. Avacharmal, and A. P. Perumal, "Mental health in the tech industry: Insights from surveys and NLP analysis," *J. Recent Trends Comput. Sci. Eng. (JRTCSE)*, vol. 10, no. 2, pp. 22–33, 2022. doi: 10.70589/JRTCSE.2022.2.3.

[10] K. Anjali, H. Negi, R. Nautiyal, and S. Bijalwan, "Psychological Mental Health Analysis using NLP and Machine Learning," in *Proc. 2024 Int. Conf. Elect. Electron. Comput. Technol. (ICEECT)*, vol. 1, pp. 1–6, Aug. 2024. doi: 10.1109/ICEECT61758.2024.10739022.