Predicting Mental Health Disorders based on Sentiment Analysis

Mr. Jadyn Dias[1]
Student,
Department of MSc. IT,
Nagindas Khandwala College,
Mumbai, Maharashtra, India
jd.jadyndias@gmail.com

Dr. Pallavi Devendra Tawde[2]
Assistant professor,
Department of BSc. IT and CS,
Nagindas Khandwala College,
Mumbai, Maharashtra, India
pallavi.tawde09@gmail.com

Abstract
The increasing prevalence of mental health disorders such as anxiety, depression, and stress calls for innovative, data-driven approaches to early detection and intervention. This research investigates the use of sentiment analysis to predict mental health disorders from social media text, utilizing machine learning techniques to enhance the accuracy and interpretability of mental health classification. We developed and evaluated multiple machine learning models, including Random Forest, Extra Trees, AdaBoost, and Logistic Regression, using a curated dataset of user-generated content reflecting various mental health states. Our rigorous experimentation demonstrates the potential of sentiment analysis combined with advanced machine learning models in identifying mental health conditions. The study highlights that while Random Forest and Extra Trees excel in prediction accuracy, Logistic Regression offers a balanced trade-off between performance and model stability, making it suitable for real-world applications. These findings contribute to the growing field of AI-driven mental health assessment, offering a foundation for integrating multimodal data, model interpretability, and ethical implementation.

Keywords: Mental health disorders, Sentiment analysis, Natural language processing, Machine learning

1. Introduction:

Mental health disorders, including anxiety, depression, and bipolar disorder, are among the leading causes of disability worldwide. These conditions not only affect the individual's emotional and psychological well-being but also have profound implications for their social and economic life. The World Health Organization (WHO) reports that depression is a significant global health issue, impacting over 264 million individuals. Similarly, Anxiety disorder impacts over 284 million people, making them one of the common mental health issues.

The symptoms of mental health disorders can vary widely but often include persistent feelings of sadness, hopelessness, anxiety, irritability, and changes in sleep and appetite. The diagnosis of mental health disorders usually relies on a mix of self-reported symptoms, clinical evaluations, and the use of standardized diagnostic instruments. However, these traditional methods can be time-consuming and may not always capture the nuances of an individual's mental state.

The proliferation of social media platforms like Facebook, Twitter, and Instagram has changed how people communicate and express themselves. These platforms provide users with the means to share their thoughts, feelings, and experiences with a wide audience. As a result, social media has become a valuable source of data for understanding human behavior and mental health. Research has shown that social media activity can reflect an individual's emotional state and even predict mental health outcomes.

While social media data offers a wealth of information, it also presents challenges. The data is often unstructured, noisy, and influenced by various factors such as cultural context, platform-specific language, and the user's intent.

However, these challenges also present opportunities for researchers to develop innovative methods for extracting meaningful insights from this data.

Sentiment analysis, also known as opinion mining, involves using natural language processing (NLP), text analysis, and computational linguistics to identify and extract subjective information from text. This technique can determine the sentiment expressed in a piece of text, categorizing it as positive, negative, or neutral. Advanced sentiment analysis techniques can interpret a broad spectrum of emotions, capturing complex emotional states such as happiness, anger, sorrow, and fear.

Sentiment analysis has been increasingly used to assess mental health through the analysis of text data from various sources, including social media posts, blogs, and online forums. For example, studies have shown that patterns of language use on social media can indicate the presence of depression, anxiety, and other mental health conditions. Sentiment analysis can help identify these patterns and provide insights into an individual's mental health, potentially leading to early intervention and support.

Research Objectives

1. To develop and evaluate multiple machine learning models for accurately predicting mental health disorders using sentiment analysis of preprocessed social media text data.
2. To classify text data into different mental health categories (e.g., Depression, Anxiety, Stress) using sentiment and linguistic features extracted from user-generated content.

2. Review of Literature:

Recent research has extensively explored the intersection of sentiment analysis, social media data, and mental health disorders. Several studies have investigated how textual data from social media platforms can be leveraged to predict mental health conditions, utilizing natural language processing (NLP) techniques to identify linguistic markers associated with various mental health issues.

Wang (2024) explored the intersection of social media data mining and mental health assessment through sentiment analysis. The study highlights how social media platforms serve as rich sources of data to assess mental health statuses by analyzing users' language, tone, and context. Leveraging machine learning algorithms and NLP tools, the research identified patterns indicative of conditions like depression, anxiety, and stress, offering real-time, scalable, and cost-effective mental health assessments. Ethical considerations, including data privacy, consent, and algorithmic biases, are also discussed as critical aspects of this interdisciplinary approach.

Uban et al. (2021) investigated the early detection of mental health disorders, including depression, anorexia, and self-harm, using social media data. The study employed deep learning models to analyze linguistic markers, emotions, and cognitive styles present in user-generated content. The researchers developed interpretable models, such as hierarchical attention networks, to capture various linguistic features and analyze their evolution over time. The findings highlight the potential of monitoring language on social media for early identification of mental disorders and assisting clinicians in diagnosis. The study also emphasizes the importance of understanding the dynamic interplay between emotions and cognitive styles in mental health assessments.

Sekulic and Strube (2020) explored the use of deep learning models to predict mental health disorders based on social media text, specifically Reddit posts. Using the Self-reported Mental Health Diagnoses (SMHD) dataset, they applied a Hierarchical Attention Network (HAN) to classify users suffering from one of nine mental disorders. The study achieved notable improvements over traditional machine learning baselines in detecting disorders like depression,

ADHD, and anxiety. They also analyzed the attention mechanism in the model to identify key phrases contributing to classification, showing correlations with previous research in mental health analysis.

Nova (2023) investigated machine learning approaches for classifying mental disorders based on textual data from Reddit. The study utilized posts from subreddits related to Borderline Personality Disorder (BPD), bipolar disorder, depression, and anxiety. After preprocessing the data by removing URLs, punctuation, and stopwords, three models—Multinomial Naive Bayes, Multi-layer Perceptron, and LightGBM—were employed. LightGBM outperformed the other models, achieving 77% accuracy when classifying based on text content. The study highlighted the potential of machine learning for automating mental health diagnosis through social media text.

Su et al. (2020) conducted a scoping review on the application of deep learning (DL) in mental health outcome research, categorizing studies into four main areas: clinical data, genetics and genomics, vocal and visual expression, and social media data. They found that DL algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), significantly improved performance in mental health diagnosis, prognosis, and risk estimation. However, challenges like data privacy and model interpretability persist, and further research is needed to bridge the gap between DL advancements and patient care.

3. Methodology:

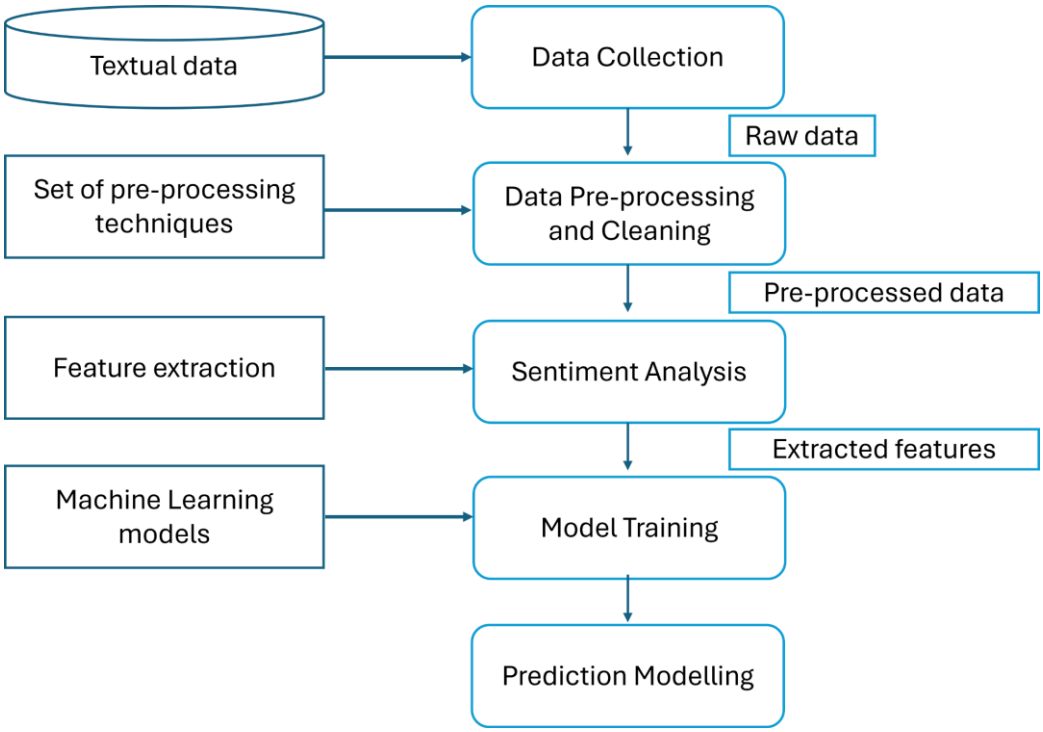

Figure 1: Methodology

3.1. Data Collection:

The dataset used in this study, titled "Sentiment Analysis for Mental Health," was sourced from Kaggle. It contains textual data reflecting various mental health states such as Anxiety, Depression, and other disorders. Each entry includes a statement and a corresponding label indicating the mental health status. The dataset was chosen for its relevance and comprehensiveness in capturing a range of sentiments associated with different mental health conditions.

3.2. Data Preprocessing and Cleaning:

The dataset underwent a series of preprocessing steps to enhance its quality and suitability for further analysis.:

- **Removing Unnecessary Columns**: Initial inspection of the dataset identified extraneous columns (e.g., 'Unnamed: 0'), which were removed to streamline the data.
- **Handling Missing Values**: Missing data entries were identified and removed to maintain the integrity of the analysis.
- **Resampling to Address Imbalance**: To address class imbalance in the dataset, a resampling function was implemented. The minority classes were upsampled to match the number of entries in the majority class using the resample function from sklearn.utils, ensuring that the model received balanced input during training.
- **Text Cleaning and Stemming**: The preprocessing step involved cleaning the text data by eliminating unwanted elements such as special characters, punctuation marks, and numerical values through the use of regular expressions. A Porter Stemmer was applied to reduce words to their root form, which helped in minimizing vocabulary size and standardizing word forms.

3.3. Feature Extraction:

- **TF-IDF Vectorization**: The cleaned text data was transformed into numerical features using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. This method helps in quantifying the importance of words within the text, allowing the model to identify key patterns associated with different mental health statuses.

3.4. Model Training:

The study employed and trained four different machine learning :

- **Random Forest Classifier**: A robust ensemble model that combines multiple decision trees to improve prediction accuracy and reduce overfitting.
- **AdaBoost Classifier**: A boosting algorithm that combines weak classifiers to create a strong classifier by focusing on misclassified instances during training.
- **Extra Trees Classifier**: An ensemble method that builds multiple decision trees using random splits of the data, providing high accuracy and efficiency.
- **Logistic Regression**: A linear model suitable for binary classification tasks, used here to handle multiclass prediction by extending it to one-vs-rest classification.

The models were trained using 80% of the dataset, with the remaining 20% reserved for testing and performance evaluation.

3.5. Model Evaluation:

- **Performance Metrics**: The trained models were evaluated using accuracy and precision scores to assess their performance in predicting mental health conditions. Both training and test accuracies were recorded to check for overfitting or underfitting.
- **Comparative Analysis**: The performance of each model was compared based on the evaluation metrics, identifying the best-performing model for predicting mental health conditions from text data.

4. Results:

The performance of four different machine learning models—Random Forest, AdaBoost, Extra Trees, and Logistic Regression—was evaluated using the test and training datasets. Below is a detailed analysis of each model's performance based on accuracy and precision metrics.

Model	Train Accuracy	Test Accuracy	Train Precision	Test Precision	Train Recall	Test Recall
Random Forest	99.98%	95.38%	99.98%	95.36%	99.97	95.18
AdaBoost	66.39%	65.98%	67.00%	66.66%	66.38	65.98
Extra Trees	99.98%	95.25%	99.98%	95.27%	99.97	95.09
Logistic Regression	90.49%	87.50%	90.36%	87.25%	90.48	87.49

Table 1: Results

The Random Forest classifier demonstrated high accuracy and precision on both training and test sets, indicating strong performance in predicting mental health statuses. However, the exceptionally high training accuracy suggests slight overfitting, where the model performs near-perfectly on the training data but slightly less well on unseen data.

The AdaBoost classifier displayed the lowest performance among the models, with moderate accuracy and precision scores on both training and test sets. The similar performance on both datasets indicates that the model has not overfit but is struggling to learn patterns from the data effectively, likely due to its sensitivity to noisy data, affecting its ability to accurately learn from the given task.

Logistic Regression showed good performance with balanced training and test accuracies and precision. This model was less prone to overfitting compared to Random Forest and Extra Trees, demonstrating more stable performance across different datasets. It achieved respectable accuracy and precision, suggesting it is well-suited for classifying mental health statuses based on textual data.

The Random Forest and Extra Trees classifiers show high accuracy and precision but are prone to overfitting, making them powerful but slightly unstable. AdaBoost has the lowest performance, suggesting it may not be suitable due to its sensitivity to noise. Logistic Regression provides a balanced performance with good generalization, making it well-suited for practical applications in predicting mental health disorders.

The research primarily aims to classify user-generated text into specific mental health categories such as Depression, Anxiety, and Stress, utilizing sentiment and linguistic features extracted from the content. For this task, the Logistic Regression model was employed due to its ability to generalize well and balance between accuracy and interpretability.

An example of the model's application is demonstrated through the function predi(), which classifies the input text **"trouble sleeping, confused mind, restless heart. All out of tune" as Anxiety.** This prediction is driven by the

identification of key phrases such as "trouble sleeping," "confused mind," and "restless heart," which are commonly associated with anxiety symptoms. The model uses these linguistic cues to estimate the probability of the text indicating anxiety, successfully categorizing the emotional state of the individual.

5. Conclusion:

This research paper examined the use of machine learning models, including Random Forest, Extra Trees, AdaBoost, and Logistic Regression, to predict mental health disorders through sentiment analysis of text data from social media. The study found that Random Forest and Extra Trees classifiers achieved the highest accuracy and precision, though their tendency to overfit emphasized the need for careful model tuning. Logistic Regression emerged as a balanced and interpretable model, making it more suitable for real-world applications, while AdaBoost struggled with the complexity of sentiment-laden mental health data.

Overall, the findings highlight the potential of machine learning and sentiment analysis as tools for early detection and intervention in mental health. While advanced models offer high performance, Logistic Regression's stability and generalization make it particularly useful for practical applications.

6. Future Scope:

The exploration of machine learning models for predicting mental health disorders based on sentiment analysis opens numerous avenues for further research and practical applications. These include integrating multimodal data, such as images and audio, to improve prediction accuracy and provide a more comprehensive understanding of mental health states. Real-time monitoring systems could enable continuous tracking and timely interventions, while personalized models can adapt to individual language use and cultural context for more accurate assessments. Addressing ethical concerns around data privacy is essential, and future studies could expand the scope to cover a wider range of mental health conditions, broadening the models' applicability.

References

[1] Wang, L. (2024). Social Media Data Mining and Mental Health Status Assessment Based on Sentiment Analysis. Journal of Electrical Systems, 20(6s), 381-386.

[2] Uban, A. S., Chulvi, B., & Rosso, P. (2021). An emotion and cognitive based analysis of mental health disorders from social media data. Future Generation Computer Systems, 124, 480-494.

[3] Sekulić, I., & Strube, M. (2020). Adapting deep learning methods for mental health prediction on social media. arXiv preprint arXiv:2003.07634.

[4] Nova, K. (2023). Machine learning approaches for automated mental disorder classification based on social media textual data. Contemporary Issues in Behavioral and Social Sciences, 7(1), 70-83.

[5] Su, C., Xu, Z., Pathak, J., & Wang, F. (2020). Deep learning in mental health outcome research: a scoping review. Translational Psychiatry, 10(1), 116.

[6] Kim, J., Lee, J., Park, E., & Han, J. (2020). A deep learning model for detecting mental illness from user content on social media. Scientific reports, 10(1), 11846.

[7] Hinduja, S., Afrin, M., Mistry, S., & Krishna, A. (2022). Machine learning-based proactive social-sensor service for mental health monitoring using twitter data. International Journal of Information Management Data Insights, 2(2), 100113.

[8] Ahmed, U., Jhaveri, R. H., Srivastava, G., & Lin, J. C. W. (2022). Explainable deep attention active learning for sentiment analytics of mental disorder. Transactions on Asian and Low-Resource Language Information Processing.

[9] Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J., & Pant, S. (2023, January). A review of machine learning and deep learning approaches on mental health diagnosis. In Healthcare (Vol. 11, No. 3, p. 285). MDPI.

[10] Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P. M. (2023). A review on sentiment analysis from social media platforms. Expert Systems with Applications, 223, 119862.