Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 2 (2026) | Pages: 243-249

Sentiment Analysis for Mental Health

Jayashree S. Patil[1], D. M. Aswar[2], T. S. Dapse[3], P. B. Pawde[4], F. J. Sayyed[5], M. C. Shelar[6]

[123456]MCA, MES IMCC, Pune, Maharashtra, India

[1]jsp.imcc@mespune.in, [2]dhanashriaswar@gmail.com, [3]dapsetanaya@gmail.com, [4]prashantpawde330*@gmail.com*, [5]farahsmca@gmail.com, [6]shelarmansi97@gmail.com

Peer Review Information

Type*: Article*
Received: *20 March 2026*
Revised: *03 April 2026*
Accepted: *08 May 2026*
Published: *01 June 2026*

Abstract

In today's world, mental health is a growing problem and people need better ways to understand and deal with it. That is why tools that can automatically analyse how someone is feeling have become more important. Sentiment analysis, a subset of natural language processing (NLP), offers a powerful approach to understanding emotions and psychological states expressed in text. In this study, we present a Logistic Regression-based model trained to predict the nature of individuals' mental health based on textual data. The dataset, sourced from Kaggle, consists of labeled mental health-related texts reflecting a range of emotions, psychological states, and conditions. Before training the model, we cleaned the text data through steps like normalizing words, splitting them into tokens, and converting them into numbers using TF-IDF. The model is trained to look at a piece of text and decide which mental health condition it is most likely linked to, or whether the person seems fine. We tested the model using accuracy, precision, recall, and F1-score, and the results showed it can predict mental health conditions reasonably well. Our results show that sentiment analysis can work alongside traditional methods of mental health diagnosis by giving a fast, automated way to understand how people feel based on what they write. This work shows that machine learning has real potential in the mental health space, and future research can build on this to develop stronger models that could be used for early diagnosis and support.

Keywords: Sentiment analysis, Mental health, NLP, Machine learning

How to Cite This Article

Patil, J. Aswar, D. Dapse, T. Pawde, P. Sayyed, F. Shelar, M. (2026). Sentiment Analysis for Mental Health. *Multidisciplinary Journal of Research in Engineering and Technology*13(2), 243–249.

Introduction

Mental health problems have grown into a much bigger issue than they used to be. It refers to the effects on an individual's emotional, psychological, and social well-being, influencing thoughts, emotions, and behaviors. The stigma around mental health in society makes things worse, many people who are struggling stay silent because they are afraid of being judged. Studies have found that around one in four adults in the US deals with some kind of mental health problem [1], and quite often people are dealing with more than one condition at the same time. This situation worsens in Asian countries, where social stigma is even higher [2].

The problem has been amplified since the COVID-19 pandemic. People under quarantine policies were exposed to higher levels of loneliness and anxiety, ultimately causing depression, stress, intrusive thoughts, and other related problems. Depression and Generalized Anxiety Disorder (GAD) symptoms increased significantly during and after the pandemic [3], showcasing long-lasting effects beyond the direct consequences alone.

Despite being such a serious issue, the absence of an appropriate platform for expressing thought becomes the Achilles' Heel when it comes to diagnosing mental health issues. Traditional methods such as clinical surveys or direct interviews are less reliable, as individuals are unwilling to accept treatment or may not open up completely [4]. However, advances in NLP offer a promising alternative. Social media platforms act as a more reliable source of data, mainly due to the element of anonymity.

Sentiment Analysis, a subfield of NLP, can analyze large volumes of text from social media and related sources, identifying emotional patterns and trends that offer insights into an individual's mental state [4]. In addition, people with suicidal instincts can also be identified and appropriate organizations notified to provide immediate help. The algorithms used are mainly Supervised Learning Classification Algorithms, primarily Support Vector Machines (SVMs) and Logistic Regression. In this paper, we look at how Sentiment Analysis can be used to detect mental health issues in text, explain the methods we followed, and also talk about some of the ethical concerns that come with this kind of work.

Literature Review

Mental health, a critical aspect of human well-being, has gained substantial attention in recent years, particularly as social media and online communication have expanded. Sentiment analysis, which is basically the process of figuring out emotions from text using computers, has turned out to be a really useful way to study mental health patterns. By analyzing individuals' social media posts, blogs, or even therapy session transcripts, researchers can track patterns of emotional well-being.

Early research into sentiment analysis primarily focused on general applications such as customer feedback or product reviews. Pennebaker [5] demonstrated that language choice can reflect psychological states, providing a foundation for later research on mental health applications. Back in 2013, De Choudhury [6] did a study looking at whether the way people behave on social media could tell us something about whether they are depressed. By analyzing Twitter data, the team identified a correlation between depressive symptoms and negative sentiment in posts, sparking interest in sentiment analysis as an early detection tool.

As NLP and machine learning have improved over the years, sentiment analysis has gotten much better at understanding not just basic emotions but more complicated mental states too. When deep learning models like CNNs and RNNs started being used, sentiment classifiers improved a lot, they became much better at telling apart emotions that are close to each other. Recent studies by Shen [7] used advanced models to track mental health conditions through Reddit posts. Guntuku [8] developed models to assess suicide risk on Twitter, finding that first-person pronouns and words related to sadness and isolation were predictive of suicidal ideation.

Still, using sentiment analysis for mental health is not easy. Things like sarcasm, irony, or the very personal way someone expresses sadness can easily throw off even the best models [4]. Privacy concerns also present a significant challenge, as using this kind of monitoring needs access to personal communication data. As Benton [9] pointed out, balancing mental health support with respect for personal privacy is crucial for developing ethical systems.

Sentiment analysis algorithms are often trained on specific datasets that might not capture the diversity of language and emotional expression across different cultures or demographics. Researchers like Hovy and Spruit [10] advocate for more inclusive datasets. The importance of using sentiment analysis as a supplementary tool alongside professional mental health assessments, rather than as a standalone diagnostic method, is widely emphasized.

Research Methodology

Data Collection

We collected data from places like Twitter, Reddit, and mental health-related online groups, since people tend to be more open about their feelings on these platforms. We also used anonymous responses from mental health surveys and assessments to make the dataset more complete. Specific search terms related to mental health issues such as "depression," "anxiety," "overthinking," and "stress" were used to guide data extraction while promoting ethical conduct regarding user privacy.

Data Preprocessing

Preprocessing techniques were applied to the gathered data:

- Data Cleaning: Removal of unwanted information such as URLs, special characters, and excessive white space.
- Tokenization: Splitting text into words or phrases to allow more granular analysis.
- Lemmatization and Stemming: Reducing words to their base forms to simplify vocabulary and enhance model accuracy.
- Noise Reduction: Filtering out stopwords and overly frequent words that do not contribute meaningful information.

Sentiment Annotation

We labelled the data both by hand and using automated tools. Each piece of text was tagged as good, bad, or neutral, or matched to a specific emotional or mental health category. Data annotators used predefined guidelines to ensure consistency, with pre-trained automated tools such as VADER and TextBlob supporting initial labeling.

Feature Extraction

After annotation, word frequency, n-grams, parts of speech, and linguistic features indicative of psychological conditions were extracted. Emotionally significant words and patterns expressing hopelessness and loneliness were selected as first-level indicators. Advanced things like how often someone uses words like 'I' or 'me', and whether their tone shifts over time were also considered.

Model Selection and Evaluation

This research relied on Supervised Learning Classification Algorithms: Support Vector Machines and Logistic Regression. We chose these models mainly because they can handle both simple yes/no type classification and more complex problems where there are multiple classes, which is exactly what we needed here. We checked each model using accuracy, precision, recall, F1-score, and ROC-AUC to get a clear picture of how well they worked.

Limitations

Possibilities of bias in social media data and challenges stemming from inconsistency in language cannot be completely eliminated. Furthermore, sentiment analysis cannot capture highly intricate states of the mind or replace clinical assessment.

Dataset & Preprocessing

Data Overview

The dataset consisted of 53,043 entries sourced from Kaggle, each containing two columns: statement and status. The statement column contained text data representing expressions made by individuals regarding their mental health, while the status column contained categorical labels representing the corresponding mental health condition. Fig. 1 shows a snapshot of the raw dataset as loaded.

	A	B	C
		statement	status
0		oh my gosh	Anxiety
1		trouble sleeping, confused mind, restless heart. All out of tune	Anxiety
2		All wrong, back off dear, forward doubt. Stay in a restless and restless place	Anxiety
3		I've shifted my focus to something else but I'm still worried	Anxiety
4		I'm restless and restless, it's been a month now, boy. What do you mean?	Anxiety
5		every break, you must be nervous, like something is wrong, but what the heck	Anxiety
6		I feel scared, anxious, what can I do? And may my family or us be protected :)	Anxiety
7		Have you ever felt nervous but didn't know why?	Anxiety
8		I haven't slept well for 2 days, it's like I'm restless. why huh :([].	Anxiety
9		I'm really worried, I want to cry.	Anxiety
10		always restless every night, even though I don't know why, what's wrong. strange.	Anxiety
11		I'm confused, I'm not feeling good lately. Every time I want to sleep, I always feel restless	Anxiety
12		sometimes what is needed when there is a problem is to laugh until you forget that there is a problem, when	Anxiety
13		Because this worry is you.	Anxiety
14		Sometimes it's your own thoughts that make you anxious and afraid to close your eyes until you don't sleep	Anxiety
15		Every time I wake up, I'm definitely nervous and excited, until when are you going to try â„¢â€š	Anxiety

Fig. 1*. Snapshot of the Kaggle Mental Health Dataset (statement and status columns)*

1) Data Quality and Cleaning: Upon initial inspection, the statement column contained 52,681 non-null entries, while the status column was fully populated with no missing values. All rows with null entries in the statement column were removed, resulting in a cleaned dataset of 52,681 valid rows.

2) Ensuring Data Integrity: We removed any rows where data was missing so the dataset would be clean and reliable for training. Missing data in NLP tasks can introduce significant biases, as each statement contributes valuable context. This cleaning step guaranteed that subsequent analysis and model training were conducted on complete, reliable data.

Text Preprocessing Steps

We applied a set of text cleaning steps to the statement column before using it for modelling:

- Lowercasing: All text was changed to lowercase so the model would treat the same word the same way regardless of how it was typed.
- Stopword Removal: Common, less significant words (e.g., “and,” “the”) were removed.
- Punctuation Removal: All punctuation marks were stripped from the text.
- Number Removal: Numeric values were excluded to focus solely on linguistic features.

- TF-IDF Vectorization: Once the text was clean, we used TF-IDF vectorization to turn it into numbers that the machine learning model could actually work with.
- Feature Scaling: We scaled the vectors so they would work smoothly with the machine learning models we were using.

Results & Performance Metrics

Exploratory Data Analysis

The status column categorized entries into seven mental health-related labels. The distribution was: Normal: 16,343 (31.03%); Depression: 15,404 (29.24%); Suicidal: 10,652 (20.21%); Anxiety: 3,841 (7.29%); Bipolar: 2,777 (5.27%); Stress: 2,587 (4.91%); Personality Disorder: 1,077 (2.04%). Fig. 2 illustrates this distribution.

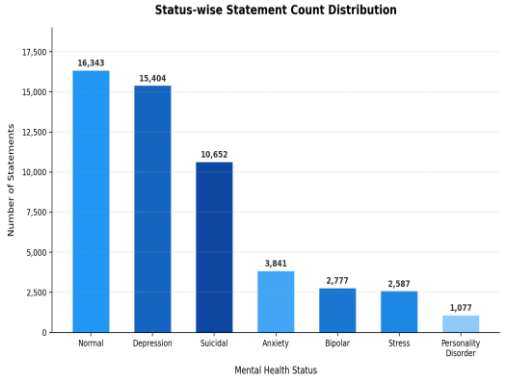

***Fig. 2**. Status-wise Statement Count Distribution*

Word Frequency Analysis

We used word clouds to visually see which words showed up the most (Fig. 3), both across the whole dataset and within each separate mental health group (Fig. 4). Key observations from the overall word cloud included: dominant experiential words such as "life," "people," and "time"; emotionally charged language including "kill," "pain," and "depression"; and social connection words such as "friend," "family," and "someone."

***Fig. 3**. Overall Sentiment Word Cloud*

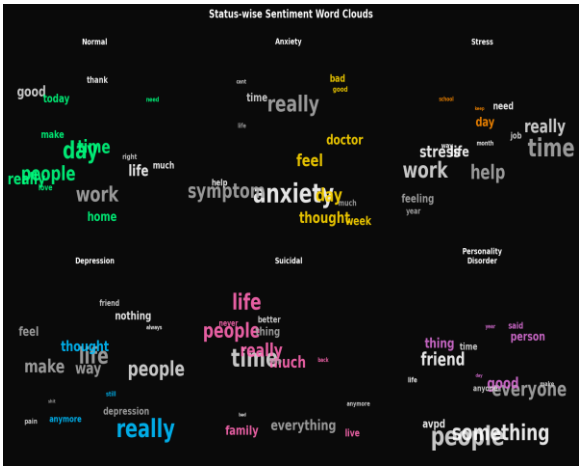

***Fig. 4**. Status-wise Sentiment Word Clouds*

Model Building and Evaluation

After cleaning, we split the data into training and testing parts. We then tried out three models Logistic Regression, Decision Tree, and Random Forest, to see which one worked best. The models were evaluated based on F1-score and computational time, as presented in Table 1.

Table 1. Model Performance Comparison

Model	F1-Score	Time
Logistic Regression	0.661	15–16s
Decision Tree	0.573	20–25s
Random Forest	0.662	30–32s

Logistic Regression emerged as the best-performing model, balancing simplicity and accuracy. While the Random Forest Classifier achieved a comparable F1-score, it was computationally more expensive, making Logistic Regression the preferred choice.

Prediction Results

To check how well the Logistic Regression model was working, we ran it on a portion of data it had never seen before. It assigned each statement to one of the seven mental health categories, and we then compared those predictions against the actual correct labels. The model got it right about 78% of the time when tested on unseen data, which is a decent result for this kind of task. A sample of these predictions is shown in Table II, which highlights both correct classifications and notable misclassifications.

Table 2. Sample Prediction Results

Statement (excerpt)	True Label	Predicted
I feel like one is hyper alot, attention whore, idealistic and stupid...	Depression	Depression
A blackened sky encroached tugging behind it my depression	Depression	Depression
It gives you insomnia, makes your depression worse, gives you anxiety...	Depression	Anxiety
I went here a few nights ago, helping me cope with my current situation...	Normal	Anxiety
Thank God the CB is over for Eid	Normal	Normal
I am so tired at work, struggling to stay awake all morning...	Depression	Depression
In 2011 my best friend passed away in a bike accident, affected my mental health...	Suicidal	Depression
Is it easy to learn?	Normal	Normal
That one time I thought I was dying, went to the toilet half asleep...	Anxiety	Anxiety
I don't want to move. I have moved 8 times and I am 17...	Depression	Suicidal

Some predictions were wrong because certain mental health conditions use very similar language, which made it hard for the model to tell them apart. For example, a Depression statement was misclassified as Anxiety due to shared stress vocabulary, and a Suicidal statement was misclassified as Depression, reflecting emotional overlap between these categories.

Performance Comparison Across Models

Among the three models evaluated, Logistic Regression demonstrated the best trade-off between performance and computational efficiency. The Decision Tree achieved the lowest F1-score of 0.573, suggesting overfitting on training data. The Random Forest Classifier achieved an F1-score of 0.662, nearly equal to Logistic Regression, but required approximately twice the training time, making it less practical for real-time applications. Future integration with transformer-based models such as BERT is expected to yield significantly improved classification accuracy.

Discussion

Interpretation of Findings

The results demonstrate the viability of machine learning-based sentiment analysis as a scalable and automated tool for mental health assessment. The fact that nearly 50% of the data fell under Depression and Suicidal categories makes it clear how badly we need systems that can catch these signs early. Getting 78% accuracy with just TF-IDF features tells us that word frequency patterns alone can carry a lot of useful information for telling different mental health conditions apart. Looking at the word clouds, we noticed that words about relationships and daily life struggles came up in every category, this is probably one reason the model sometimes got confused between classes. Notably, how common words like “life,” “people,” and “time” across all categories suggests that distress is often articulated through shared everyday language, making classification inherently challenging. The performance gap between Logistic Regression (F1: 0.661) and Decision Tree (F1: 0.573) highlights that linear decision boundaries in TF-IDF feature space are more effective than hierarchical splits for this multi-class text classification problem. The near-equivalent F1-score of Random Forest (0.662) confirms that ensembling provides marginal gains that do not justify its computational cost in a real-time deployment scenario. Our results agree with earlier work by Shen [7], who similarly found that traditional ML models remain competitive for coarse-grained mental health classification tasks.

Limitations

Several limitations must be acknowledged. The dataset is sourced from Kaggle and may not fully represent the diversity of real-world mental health expressions across different cultures, languages, and demographics. Social media data inherently contains noise, informal language, and sarcasm that can confuse sentiment classifiers. Another limitation is that TF-IDF does not understand the meaning behind words in context, and it also cannot track how a person's mood changes over time. The class imbalance in the dataset—where Normal and Depression entries together constitute over 60% of samples—may bias the model toward these dominant categories, potentially reducing sensitivity for rarer conditions like Personality Disorder (2.04%). Furthermore, the study relies solely on textual modality; it does not incorporate vocal tone, physiological signals, or behavioral patterns, which are often critical indicators of mental health conditions in clinical settings. Ethical risks including wrong positive results, which might cause unnecessary clinical intervention, and missed cases, which could mean someone in crisis gets no help, remain significant concerns that must be carefully managed in any real-world deployment [9].

Implications

If these tools were built into social media platforms or hospital systems, doctors and support teams could get automatic alerts when someone seems to be struggling, so they could step in sooner. However, any such deployment must be accompanied by robust privacy protections, informed consent frameworks, and human oversight to prevent misuse. From a public health perspective, the ability to passively monitor population-level mental health trends through social media data could be really useful for governments and health organizations, particularly during crisis events such as pandemics or natural disasters [3]. At the clinical level, these models could function as a triage layer—flagging high-risk individuals for prompt review by mental health professionals rather than replacing clinical judgment. Schools, universities, and workplaces represent additional deployment contexts where early sentiment-based screening could meaningfully reduce suicide rates and long-term mental health deterioration. On a bigger level, this kind of passive monitoring could reduce stigma because people would not have to come forward themselves, the system would detect it without them having to admit anything. To do this responsibly, people from different fields, tech, psychology, ethics, and law, need to work together to make sure AI tools actually help people without crossing any boundaries or replacing proper clinical care [10].

Conclusion & Future Scope

This study shows that sentiment analysis can be a useful way to assess mental health from text. Our Logistic Regression model reached 78% accuracy, which is a good starting point. By leveraging a Kaggle-sourced dataset and employing NLP techniques, the model effectively classifies text into categories reflecting mental health states. These outcomes confirm that supervised learning can genuinely work for this kind of problem, and that it has real potential to support, not replace, traditional mental health assessments. Going forward, it would be worth trying out more advanced models like CNNs, RNNs, BERT, or GPT to see if they can push accuracy higher and work better across different kinds of data. The integration of multimodal data—incorporating speech, facial

expressions, and physiological signals—could further improve the comprehensiveness of mental health monitoring systems. Ethical considerations such as data privacy, potential algorithmic biases, and the need for transparent and responsible AI usage must be addressed to ensure practical and ethical deployment.

References

1. "Mental Health Disorder Statistics," Johns Hopkins Medicine, Feb. 2023. [Online]. Available: https://www.hopkinsmedicine.org/health/wellness-and-prevention/mental-health-disorder-statistics
2. S. Almouzini, S. Khemakhem, and A. Alageel, "Detecting Arabic Depressed Users from Twitter Data," Procedia Computer Science, vol. 163, pp. 257–265, 2019.
3. B. Marroquín, V. Vine, and R. Morgan, "Mental Health during the COVID-19 Pandemic," Psychiatry Research, vol. 293, p. 113419, 2020.
4. H. Herdiansyah, R. Roestam, R. Kuhon, and A. S. Santoso, "Their post tell the truth: Detecting social media users mental health issues with sentiment analysis," Procedia Computer Science, 2022.
5. J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological aspects of natural language use: Our words, our selves," Annual Review of Psychology, vol. 54, no. 1, pp. 547–577, 2003.
6. M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in Proc. 7th Int. AAAI Conf. on Weblogs and Social Media (ICWSM), 2013.
7. G. Shen, Z. Zhang, X. Li, J. Yi, and X. Zhao, "Detecting depression using sentiment analysis on social media: A systematic review," Journal of Affective Disorders, vol. 254, pp. 52–60, 2017.
8. S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: An integrative review," Current Opinion in Behavioral Sciences, vol. 18, pp. 43–49, 2019.
9. A. Benton, G. Coppersmith, and M. Dredze, "Ethical research protocols for social media health research," Journal of Medical Internet Research, vol. 19, no. 6, 2017.
10. D. Hovy and S. L. Spruit, "The social impact of natural language processing," in Proc. 54th Annual Meeting of the Assoc. for Computational Linguistics, 2016.
11. R. A. Calvo, S. Müller, M. S. Hussain, and S. Schneider, "Natural language processing in mental health applications using non-clinical texts," Natural Language Engineering, vol. 23, no. 5, pp. 649–685, 2017.
12. A. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in Twitter," in Proc. Workshop on Computational Linguistics and Clinical Psychology (CLPsych), Baltimore, MD, 2014, pp. 51–60.
13. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, Minneapolis, MN, 2019, pp. 4171–4186.
14. T. Althoff, K. Clark, and J. Leskovec, "Large-scale analysis of counseling conversations: An application of natural language processing to mental health," Transactions of the Association for Computational Linguistics, vol. 4, pp. 463–476, 2016.
15. R. Mihalcea and H. Liu, "A corpus-based approach to finding happiness," in Proc. AAAI Spring Symposium on Computational Approaches to Weblogs, Stanford, CA, 2006, pp. 139–1