# *Sentiment Analysis for Depression using twitter comments*


## *Abstract—*

It is a chronic mental disorder called depression that remains unnoticed until the tragically worst hours. It seems people have developed the art of textually expressing their emotions with increasing avenues for online communication. That, thus sentiment analysis could be utilized for recognizing the behavior pattern of depression. Many machine learning models were applied in this work for classifying the user comments on different online platforms as depressive and non-depressive [2]. A user comment sample underwent preprocessing methods like tokenization, stop-word filtering, and TF-IDF embedding [19]. Random forest came out to be the most viable due to the compromise between hardware utilization and accuracy after testing several algorithms. Performance verification was done based on the reduced metrics like accuracy, precision, recall, and F1-score. The research propounds a future AI-based mental health monitoring tool which paves the way for its integration with real-timed platforms and deep neural networks [4].


## *1.Introduction*

The mental-illness-related disorders, nearly all of which are classified as mental-related, have their problems manifested mostly in depression, which could be rated as one of the most prominent issues facing health across the globe at this age. Most of the people skip these early symptoms mostly, considering that self-report questionnaires and a conventional clinical evaluation fail to address these aspects of the disease. The free forums accessed via the internet have allowed everyone into speaking freely through users' comments, which have now become a former goldmine for early mental analysis [4].

This is to create a cheaper but scalable sentiment analysis model that would learn to classify comments as depressive or non-depressive from machine learning. It thereby bridges the gap in diagnosis from the traditional methods to the machine by processing emotional expressions in text through Natural Language Processing (NLP) and machine learning. Many different types were recorded and through several experiments regarding their efficiency [6]: SVM, Naive Bayes, Neural Networks, etc. However, Random Forest was considered the most effective due to its relatively high accuracy, efficiencies with available computing power, and ease of interpretation.

They gather publicly available datasets of user reviews under the category of user reviews associated with depression and perform pre-processing, which includes tokenization, stop-word filtering, lemmatization, and TF-IDF for feature extraction [19]. The data in turn is trained and tested using the specified performance measures with a Random Forest Model. The area of this work is toward developing AI-based monitoring tools related to mental health, and it encompasses possible future directions for research concerning the integration of real-time social media and further improvement through deep learning techniques [1].

Future sections of this paper will consist of related works, literature survey, preparation of datasets, research methodology, result discussions, and future improvement prospects. This research on machine learning application in sentiment analysis expects to develop AI-based tools for preventive mental health intervention and support.

## 2. Problem Statement

*Depression is a current and relevant mental health issue that is often ignored by all until it has reached the verge of a full-blown crisis. Written discourse, especially through social media, may carry the subtle expressions of mental promotion in articulating sadness, disappointment, or emotional discomfort. However, it is often difficult to detect such cues because of:*

- Emotional Expression Has Various Outlets: Many persons have their own ways of expressing different emotional feelings. In turn, this tends to make the detection of emotions less standardizable.
- Insufficient Number of Labeled Data: There are few high-quality labeled datasets for depressive sentiment analysis and hence this directly obstructs model training.
- Ethical and Privacy Concerns: The greatest degree of ethical soundness and anonymity warrants consideration since mental health information deals with several sensitive issues.

*This research shall overcome these challenges by developing an accurate, ethical, scalable system that detects depressive sentiments amidst text, applying sophisticated machine learning techniques.*

### A. Why

- *Transformation potential by catching early signs of depression.*
- *Knowledgeable Awareness and Prevention: Prepares the individual to recognize the earliest indicators of mental health problems and do something about it.*
- *Support for the Professionals: Assists mental health professionals in better identifying those at risk.*
- *Positive Digital Spaces: Enables the platforms to identify harmful trends for prevention and promotion of healthful digital behaviors. It, through text analysis, is designed to work as a preventive mental health monitoring and intervention system.*

### B. How

*The project follows a step-wise procedure to ensure precision, reliability, and ethical compliance with the following steps:*

- *Problem Definition: Set this task into a binary classification-either classify the textual work an or not depressive one within it.*
- *Data Collection: Source collection of text data from net available sites e.g. from social media avenues while considering anonymity-compliance for ethical purposes.*
- *Data Preprocessing:*
  - *Clean up the text data*
  - *Normalize text*
  - *Word or text tokenization*
  - *Stop words and stemming/lemmatization removal.*
  - *Sample balancing into equal size of depressive part and nondepressive parts.*
- *Feature Engineering: Convert text to some numeric representation using techniques like TF-IDF, or also use more recent embedding techniques like BERT.*

- *Model Building: Develop a Random Forest classifier. Tune hyperparameters to fetch good performance by counting the number of trees and also setting the depth for trees.*
- *Testing of Model: Verify the correctness of the system using accuracy, precision, and recall measures alongside an F1 score.*
- *Deployment: Develop an application user friendly which analyses input text and gives output for real-time predictions.*
- *Continual improvement: System will be improved through real-time testing, user feedback, and increments.*

## 3. Design

Some main characteristics incorporated into the system are:

- High Accuracy- Trustworthy in the recognition of depressive sentiments.
- Robustness- Capability of handling the noisy and diversely effective data.
- Interpretability- Indicates significant linguistic features associated with depression.
- Scalability- This system efficiently processes large volumes of text data.

## 4. Methodology

The present study is about the analysis of sentiment in Twitter messages and the detection of negative sentiments primarily through machine learning approaches. Out of the many models tested, Random Forest proved to be the best in terms of efficiency and accuracy, and is therefore taken as the core classifier for this approach.

### A) Dataset and Preprocessing

- This dataset is composed of labeled tweets, which come from the file "twitter_training.csv". Each tweet has its specific sentiment attached to it: Positive, Negative, Neutral, and Irrelevant. In the binary classification, tweets bearing the "Negative" label were encoded with 1, and the others encoded with 0.
- The process of cleaning and normalizing the text constituted the preprocessing steps that included all the steps depending on URL removals, mentions, hashtags, and non-alphabetic characters were deleted using several regular expressions. The entire text was minimized. Then, some basic NLP processing was carried out: tokenization, stopword removal using the NLTK stopword corpus, and lemmatization using the WordNetLemmatizer to convert the words into base forms, thus allowing the model to concentrate on sentiment and sufficiently reduce the noise.

### B) Feature Extraction

*After cleaning the tweets, they were then transformed numerically via Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. Constraining the maximum dimensions at 5,000 to ensure that meaningful context is retained, these features were ultimately used as input to the machine learning model.*

### C) Model Selection and Justification

- Modules such as BERT and Neural Networks were included among the classification models in the study, along with Naive Bayes, Logistic Regression, SVM, and XGBoost. They yield great results, however, from their training, which is quite expensive in terms of resources and time, and BERT can merely score 73.7% even in a high-end complex setup.
- Random Forest, on the other hand, gave an accuracy of 93.5%, besides cost-effective. It gives an advantage by being able to train in few seconds and free from GPU or even deep learning framework, as it has been into scikit-learn itself. It will also perform well in dealing with nonlinear relationships using the ensemble-based method in text data while providing transparency along with understanding through feature importance analysis as an interpretation of the model behavior.

### *D ) Training and Evaluation*

Data were split into training and test sets in an 80/20 ratio. RandomForestClassifier with 50 estimators was fitted on the data for training. The model evaluation metrics were Accuracy, Precision, Recall, F1 score, and Confusion matrix; these showed that the Random Forest model exhibited better performance in terms of accuracy and computational efficiency compared to the rest of the classifiers.

### *E ) Conclusions*

Random Forest is really very robust and thus a sound approach for intelligible sentiment classification. The performance metrics it offers are not much away from that of neural networks and yet it takes far less resources to implement. Thus, this is suitable for real-life usage where interpretability and speed hold better consideration than many others because of simplicity coupled with maximum scalability.

## *5. Algorithm Of Choice*

It is because of these that Random Forest was ranked as the primary model for the sentiment analysis project. Random Forest hits a balance of accuracy, efficiency, and interpretability. Compared to its other contenders-BERT, Neural Networks, Naive Bayes, SVM, and XGBoost-Random Forest received an accuracy of 93.5%. This modeling was comparable to the delivered performances of other models-neural networks obtained a noted performance of 93.3% while BERT had 73.7%-as it was found to be much cheaper on computing resource costs.

Random Forest stands out now from BERT or deep neural networks in not demanding GPU or high-memory settings. Random Forest is therefore most suitable under a resource-constrained environment, like a laptop: lightweight, training in seconds, and performs almost as well on low-spec hardware. Its seamless integration with scikit-learn also avoids all the complexity related to the development pipeline that would have been needed had we gone with GPUs and other frameworks like TensorFlow or PyTorch.

Yet another benefit is interpretability. Specific importance scores associated with features of a Random Forest can help in identifying words that are contributing most to the prediction of sentiment. Thus, an increased level of transparency and ease of debugging as compared to black-box models like BERT can be conferred on it.

Thus, Random Forest selected because it is suitable for high accuracy while requiring low computational costs; it makes fast training and inference possible, integrates well with current tools, and provides very clear model interpretability. All of which make this model competent and efficient for performing sentiment analysis on Twitter data.

## *6. Literature Review*

| Authors | Title | Algorithms Used | Models Used | Dataset | Accuracy | Findings | Limitations |
|---|---|---|---|---|---|---|---|
| Vandana, et al (2023) | A hybrid model for depression detection using deep learning | CNN, LSTM, Bi-LSTM | Textual CNN, Audio CNN, Hybrid LSTM/Bi-LSTM | DAIC-WOZ Database (189 sessions) | Textual CNN: 92 Audio CNN: 98 Bi-LSTM: 88% | Audio > Text (98% vs 92%). Bi-LSTM > LSTM. Hybrids = more robust. | Imbalanced data (4:1), Bi-LSTM slow (5+ hrs), limited to DAIC-WOZ. |
| Lamia Bendebane et al (2023) | *A Multi-Class Deep Learning Approach for Early Detection of Depressive and Anxiety Disorders Using Twitter Data* | Deep Learning (CNN, RNN, LSTM, GRU, BiRNN, BiLSTM, BiGRU), Grid Search | Hybrid models (e.g., CNN-BiGRU, CNN-BiLSTM) | 3.17M tweets (English) | 93.38 (CNN-BiGRU) | Multi-class > Binary<br>Detects depression vs. anxiety well<br>Grid search tuned learning rate | Labeling issues<br>Not tested on non-English tweets<br>Needs clinical validation |
| Nikhil Goel et al (2024) | *Automated Depression Detection System: Integrating Sentiment Analysis and Behavioural Data* | | Hybrid (SVM + Decision Trees), Neural Networks | 1,000 subjects (text + wearable data) | 90 (Hybrid Model) | Sentiment (85%) + behavior (r=0.7)<br>Hybrid model: F1 = 0.89<br>30% symptom reduction post-intervention | Reliance on self-reported data (bias risk).<br>Contextual ambiguity in sentiment analysis.<br>Device variability affects behavioural data quality. |
| Shumaila Aleem et al (2022) | *Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions* | SVM, RF, KNN, DT, AdaBoost, CNN, LSTM, DCNN, XGBoost | Classification, Deep Learning, Ensemble | EEG, social media (Twitter, Reddit), clinical records (PHQ-9, BDI-II) | 76.6–98.32 | SVM and RF are robust; EEG-based DL models achieve high accuracy; multimodal approaches show promise. | Small sample sizes, lack of standardized datasets, limited clinical applicability. |
| Mumtaz & Qayyum (2019) | *EEG-based DL model for diagnosing unipolar depression* | 1DCNN, LSTM | Deep Learning | EEG (30 healthy, 33 MDD) | 98.32 | High accuracy in EEG-based depression classification | High accuracy in EEG-based depression classification |
| Faye Beatriz Turnaliuan et al (2024) | *Development of a two-stage depression* | LSTM with Dropout, GRU, CNN, | Two-stage model (Binary + multi-class) | 86,163 tweets (English/Filipino) annotated | Stage 1: 91 (F1: 0.90) | Two-stage model: Binary detection + 6 | Errors from word associations, |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *symptom detection model: application of neural networks to twitter data* | Naïve Bayes, Random Forest | | with 13 depression categories | Stage 2: 83 (F1: 0.81) | symptom types<br>LSTM + Dropout = best performance | negation, imbalance<br>Limited to English/Filipino; excludes regional languages |
| Rafael Salas-Zárate (2022) | Detecting Depression Signs on Social Media: A Systematic Literature Review | SVM, Logistic Regression, Neural Networks, Random Forests | Word Embedding, N-grams, Bag of Words, Tokenization | Twitter, Reddit, Facebook, Instagram, Weibo, NHANES | N/A | Twitter + SVM/embeddings most used<br>Python tools, cross-validation standard | Limited to studies from 2016-2021. Focused mainly on English-language platforms. |
| Arora and Arora (2019) | Mining Twitter Data for Depression Detection | SVM, Decision Trees | N-grams, Bag of Words, Stemming | Twitter (3754 tweets) | N/A | Compared SVM and Decision Trees for depression detection. Found SVM to be more effective. | Small dataset. Limited to Twitter. |
| Nadeem (2016) | Identifying Depression on Twitter | SVM, Neural Networks | Bag of Words, TF-IDF | Twitter (1,253,594 tweets) | N/A | Used TF-IDF and Bag of Words for feature extraction. Compared SVM and Neural Networks. | Large dataset but limited to Twitter. No accuracy reported. |
| Yazdavar (2020) | Multimodal Mental Health Analysis in Social Media | SVM, Neural Networks | Word Embedding, LIWC, Cohen's Kappa | Twitter (8770 users) | N/A | Combined linguistic and behavioural features for depression detection. Used Cohen's Kappa for validation. | Complex multimodal approach may not be scalable. |
| Chiong (2021) | A Textual-Based Featuring Approach for Depression Detection Using Machine Learning Classifiers and Social Media Texts | SVM, Neural Networks | N-grams, Bag of Words | Twitter, Facebook (22,191 records) | N/A | Compared SVM and Neural Networks for textual analysis. Found SVM to perform better. | Limited to textual features. No accuracy metrics. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Katchapakirin (2018) | Facebook Social Media for Depression Detection in the Thai Community | SVM | LIWC, RapidMiner | Facebook (35 users) | N/A | Developed a depression detection algorithm for Thai-language Facebook posts. Used LIWC for feature extraction. | Small dataset. Limited to Thai language |
| Wongkoblap (2019) | Predicting Social Network Users with Depression from Simulated Temporal Data | Neural Networks | Word Embedding, Softmax Function | Simulated data | N/A | Used temporal data and word embedding for depression prediction. Applied Softmax for classification. | Simulated data may not reflect real-world scenarios. |
| Bazen Gashaw Teferra (2024) | Screening for Depression Using Natural Language Processing: Literature Review | SVM, Logistic Regression, Neural Networks, Transformers (BERT, GPT) | Sentiment Analysis, Linguistic Markers, Word Embeddings, LLMs | DAIC-WOZ, Weibo, Twitter, Reddit | 82.3 - 91 | NLP (sentiment, LLMs) = high accuracy<br>Key issues: ethics, cultural sensitivity | Limited databases, no meta-analysis<br><br>English/Chinese focus limits generalizability |
| Rathners (2017) | How did you like 2017? Detection of language markers of depression and narcissism | Logistic Regression | LIWC-based features | Personal narratives (220 participants) | $R^2 = 0.104$ | Demonstrated the use of LIWC for detecting linguistic markers of depression in personal narratives. | Small dataset; limited to specific narrative context. |
| Prabhu (2022) | Harnessing emotions for depression detection | LSTM | Word2vec | DAIC-WOZ (189 sessions) | 82.3 | High accuracy: LSTM + Word2Vec for emotion-based detection | Clinical data only; may not generalize to social media |
| Islam (2018) | Depression detection from social network data using machine learning techniques | Decision Tree | LIWC | Facebook comments (7145 comments) | F-measure = 0.71 | Used decision trees with LIWC features to detect depression in Facebook comments. | Focused on Facebook; may not apply to other platforms. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Choudhury (2021) | Predicting depression via social media | SVM | LIWC (22 linguistic styles) | Twitter (554 users) | 72.4 | Identified linguistic styles associated with depression on Twitter using SVM. | Limited to Twitter; potential bias in user selection. |
| Stankevic (2018) | Feature engineering for depression detection in social media | SVM | Word Embeddings | CLEF and eRisk 2017 (887 users) | $F_1$-score = 63.4 | Word embeddings + SVM used; feature engineering was key | Moderate performance; dataset limitations. |
| Lopez-Otero (2017) | Depression detection using automatic transcriptions of de-identified speech | SVM | GloVe | DAIC-WOZ (189 sessions) | $F_1$-score = 73 | Applied GloVe embeddings with SVM for speech-based depression detection. | Small dataset; limited to clinical settings. |
| Mallol-Ragolta (2019) | A hierarchical attention network-based approach for depression detection | Hierarchical Attention Network | GloVe | DAIC-WOZ (189 sessions) | UAR = 0.66 | Proposed a hierarchical attention network for depression detection using GloVe embeddings | Complex model; requires large datasets for training |
| Dinkel (2020) | Text-based depression detection on sparse data | SVM | ELMo | DAIC-WOZ (189 sessions) | $F_1$-score = 84 | Achieved high performance using ELMo embeddings for sparse data depression detection. | Limited to specific datasets; may not generalize. |
| Rutowski (2022) | Depression and anxiety prediction using deep language models and transfer learning | Transformers | GloVe | American English spontaneous speech (16,000 sessions) | AUC = 0.8 | Transformers used for prediction; transfer learning proved effective | Focused on English speech; may not apply to text data. |
| Korti (2022) | Depression detection from Twitter posts using NLP and | LSTM | Word Embeddings | Twitter | 91 | Achieved high accuracy with LSTM for Twitter-based | Limited to Twitter; potential bias in data collection. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | machine | | | | | depression | |
| Tejaswini (2024) | Depression detection from social media text analysis using hybrid deep learning | FastText + LSTM | FastText | Reddit and Twitter (13,000 posts) | 87 | Combined FastText and LSTM for high-accuracy depression detection on social media. | Focused on English platforms; may not generalize to other languages. |
| Senn (2022) | Ensembles of BERT for depression classification | BERT | Transformers | DAIC-WOZ (189 sessions) | $F_1$-score = 0.62 | BERT ensembles used for depression classification in clinical interviews | Small dataset; computationa l complexity. |
| Hayati (2022) | Depression detection on Malay dialects using GPT-3 | GPT-3 | Few-shot Learning | Interview questions (53 participants) | $F_1$-score = 0.64 | Applied GPT-3 for few-shot learning in Malay dialect depression detection. | Small dataset; limited to specific cultural context. |
| Németh (2022) | Bio, psycho, or social: supervised machine learning to classify discursive framing | DistilBERT | Transformers | SentiOne (80,000 posts | 73 | Used DistilBERT to classify discursive framing of depression in online health communities | Focused on discursive framing; not direct depression detection. |

# *7. Software and Hardware requirements*

## *A ) Software*

- *Python (for Machine Learning & NLP)*
- *Scikit-learn: Machine Learning library*
- *Pandas & NumPy: Data manipulation*
- *TensorFlow / PyTorch (Optional): Deep learning support*

## *B ) Hardware*

- *Minimum Requirements*
  - *Processor: Intel Core i5 (8th Gen) / AMD Ryzen 5*
  - *RAM: 8 GB*
  - *Storage: 256 GB SSD*
  - *GPU: Integrated Graphics*
- *Recommended Requirements*
  - *Processor: Intel Core i7 (10th Gen) / AMD Ryzen 7*
  - *RAM: 16 GB or more*

        - *Storage: 512 GB SSD or higher*
        - *GPU: NVIDIA RTX 3060 / AMD Radeon equivalent (for deep learning models)*
    - *Additional Requirements*
        - *Stable internet connection*
        - *External storage (for dataset backup)*
        - *Cloud Computing (Google Colab, AWS, or Azure for large-scale processing)*

## *8. Future Scope*

- *Live- Event: Their being integrated in social media platforms is the main aim of the project to facilitate real-time textual analysis.*
- *State-of-the-art Models: Employing along with the advanced models such as BERT is tried to supplement the analyzer for contextual perception.*
- *Expert Collaboration: Collaboration with mental health specialists to fine-tune tool accuracy and applicability.*
- *The success of this system keeps hope alive to scale this into a multilingual analysis and thus broaden the global reach of the project.*

## *9. Conclusions*

This project, by virtue of Random Forest algorithms, endeavors to establish dependable and scalable tools for sentiment analysis of depression. The monitoring of this system takes a forward-looking view towards mental health which allows people and professionals to identify the disease in its early stages through textual data. Reliability, interpretability, and ease of use are thus maximally assured, providing a bit of support through scientific mechanisms against mental health hurdles.

## *10. References*

[1] Vandana, et al (2023) ; A hybrid model for depression detection using deep learning

[2] Lamia Bendebane et al (2023) ; A Multi-Class Deep Learning Approach for Early Detection of Depressive and Anxiety Disorders Using Twitter Data

[3] Nikhil Goel et al (2024) ; Automated Depression Detection System: Integrating Sentiment Analysis and Behavioural Data

[4] Shumaila Aleem et al (2022) ; Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions

[5] Mumtaz & Qayyum (2019) ; EEG-based DL model for diagnosing unipolar depression

[6] Faye Beatriz Turnaliuan et al (2024) ; Development of a two-stage depression symptom detection model: application of neural networks to twitter data

[7] Rafael Salas-Zárate (2022) ; Detecting Depression Signs on Social Media: A Systematic Literature Review

[8] Arora and Arora (2019) ; Mining Twitter Data for Depression Detection

[9] Nadeem (2016) ; Identifying Depression on Twitter

[10] Yazdavar (2020) ; Multimodal Mental Health Analysis in Social Media

[11] Chiong (2021) ; A Textual-Based Featuring Approach for Depression Detection Using Machine Learning Classifiers and Social Media Texts

[12] Katchapakirin (2018) ; Facebook Social Media for Depression Detection in the Thai Community

[13] Wongkoblap (2019) ; Predicting Social Network Users with Depression from Simulated Temporal Data

[14] Bazen Gashaw Teferra (2024) ; Screening for Depression Using Natural Language Processing: Literature Review

[15] Rathners (2017) ; How did you like 2017? Detection of language markers of depression and narcissism

[16] Prabhu (2022) ; Harnessing emotions for depression detection

[17] Islam (2018) ; Depression detection from social network data using machine learning techniques

[18] Choudhury (2021) ; Predicting depression via social media

[19] Stankevic (2018) ; Feature engineering for depression detection in social media

[20] Lopez-Otero (2017) ; Depression detection using automatic transcriptions of de-identified speech

[21] Mallol-Ragolta (2019) ; A hierarchical attention network-based approach for depression detection

[22] Dinkel (2020) ; Text-based depression detection on sparse data

[23] Rutowski (2022) ; Depression and anxiety prediction using deep language models and transfer learning

[24] Korti (2022) ; Depression detection from Twitter posts using NLP and machine learning techniques

[25] Tejaswini (2024) ; Depression detection from social media text analysis using hybrid deep learning

[26] Senn (2022) ; Ensembles of BERT for depression classification

[27] Hayati (2022) ; Depression detection on Malay dialects using GPT-3

[28] Németh (2022) ; Bio, psycho, or social: supervised machine learning to classify discursive framing