

DeepSentiMind: An Ensemble Intelligence Framework for Social Media-Based Mental Health Prediction

Umaira Shahneen¹, Dr. Anand N Patil²

¹ Student, Department of Artificial Intelligence & Data Science, Sharnbasva University, Kalaburagi, Karnataka, India, umairashahneenkhan@gmail.com

² Professor, Department of Artificial Intelligence & Data Science, Sharnbasva University, Kalaburagi, Karnataka, India, patilanand1990@gmail.com

ABSTRACT

Sentiment analysis plays a vital role in identifying emotional patterns expressed through social media content, enabling early recognition of mental health concerns. DeepSentiMind presents an intelligent ensemble framework for social media-based mental health prediction using natural language processing and machine learning. The framework preprocesses textual data, extracts meaningful linguistic features, and combines Support Vector Machine, Convolutional Neural Network, and Logistic Regression through soft voting to improve classification reliability. Publicly available Reddit datasets support experimental evaluation, while TF-IDF representations and pretrained GloVe embeddings strengthen textual understanding. Performance comparison demonstrates superior predictive capability, achieving accuracy exceeding ninety-five percent compared with individual classifiers. A Django-based web application provides real-time prediction, user interaction, dataset management, model retraining, and visualization dashboards for performance monitoring. The proposed approach supports timely identification of depression indicators from user-generated posts while maintaining a scalable architecture suitable for research and practical screening environments. Experimental findings indicate that ensemble learning reduces individual model limitations and improves prediction consistency. The system offers separate interfaces for users and administrators, enabling secure data handling, historical result tracking, and accuracy visualization. Future enhancements include multilingual support, transformer-based models, multimodal analysis, continuous monitoring, and clinical validation to improve robustness, generalization, usability, reliability, deployment readiness, future research.

Keywords: Sentiment Analysis, Mental Health Prediction, Social Media Analytics, Natural Language Processing, Ensemble Learning, Support Vector Machine, Convolutional Neural Network, Logistic Regression, Depression Detection, Django Framework.

I. INTRODUCTION

Mental health has become one of the most significant public health concerns worldwide because psychological disorders affect individuals regardless of age, profession, or social background. Conditions such as depression, anxiety, and emotional distress influence daily activities, interpersonal relationships, academic performance, and workplace productivity. Early recognition of these conditions is essential because timely intervention can reduce their long-term impact and improve overall quality of life. Conventional mental health assessment primarily depends on clinical interviews, psychological questionnaires, and professional evaluation. Although these approaches remain reliable, they are generally reactive, time consuming, and dependent on individuals seeking medical assistance. The increasing use of digital communication platforms has created new opportunities to analyze publicly shared textual information for identifying behavioral and emotional patterns associated with mental health conditions [1]. Social media platforms have evolved into major communication channels where millions of users regularly express opinions, emotions, experiences, and personal challenges through textual posts. These platforms generate large volumes of user-generated content that reflects everyday thoughts and emotional states. Unlike traditional surveys, social media provides continuous and naturally occurring data that can support computational analysis without requiring structured interviews. Recent advances in artificial intelligence and natural language processing have enabled automated systems to process this textual information efficiently. Sentiment analysis has emerged as an important technique for identifying positive, negative, and neutral emotional expressions within social media posts, providing valuable insights into psychological well-being and supporting early detection of potential mental health issues [2]. Natural Language Processing (NLP) enables computers to understand, preprocess, and interpret human language by removing noise, extracting linguistic features, and

representing textual information in a machine-readable format. Machine learning algorithms can then classify social media posts based on learned behavioral patterns and emotional characteristics. Traditional approaches such as Support Vector Machine (SVM) and Logistic Regression have demonstrated strong performance in text classification because of their effectiveness in handling high-dimensional textual features. Deep learning techniques, particularly Convolutional Neural Networks (CNN), further improve performance by automatically learning semantic relationships and contextual patterns within textual sequences. Despite these advancements, individual classifiers often exhibit limitations related to generalization, prediction consistency, and sensitivity to diverse writing styles across different users [3].

To overcome these challenges, ensemble learning has become an effective strategy for combining multiple classification models into a unified prediction framework. Instead of depending on a single algorithm, ensemble methods integrate complementary strengths from different classifiers to improve overall prediction accuracy, reduce variance, and minimize classification errors. The combination of SVM, CNN, and Logistic Regression through soft voting enables balanced decision making by considering probabilistic outputs from each individual classifier. Such an approach provides greater robustness when analyzing complex social media text containing informal language, abbreviations, spelling variations, and diverse emotional expressions commonly observed across online communities [4]. The proposed study introduces DeepSentiMind, an intelligent ensemble framework designed for social media-based mental health prediction using natural language processing and machine learning techniques. The framework performs text preprocessing, feature extraction using TF-IDF representations and pretrained GloVe embeddings, followed by classification through an ensemble voting mechanism integrating SVM, CNN, and Logistic Regression models. Publicly available Reddit datasets containing depression-related and non-depression posts are utilized for training and evaluation. A Django-based web application enables users to submit textual content for real-time prediction, while administrators can manage datasets, retrain models, and visualize classification performance through interactive dashboards. The modular architecture supports efficient deployment and practical usability within research-oriented environments [5]. Experimental evaluation demonstrates that the ensemble framework achieves higher predictive accuracy than individual machine learning models, indicating improved reliability for early mental health prediction. The integration of multiple classifiers reduces individual model weaknesses while maintaining consistent performance across diverse textual inputs. Although the current framework focuses primarily on English-language social media posts, future enhancements may incorporate multilingual support, transformer-based language models, multimodal analysis involving images and audio, continuous monitoring, and clinical validation to improve scalability and real-world applicability. Such developments can strengthen computational mental health assessment and contribute to timely identification of individuals who may require professional psychological support while preserving an efficient, intelligent, and scalable prediction framework [6], [7].

II. RELATED WORKS

Early research on computational mental health prediction established the foundation for analyzing psychological conditions through social media content. One of the pioneering contributions was presented by Coppersmith et al., who demonstrated that linguistic patterns extracted from Twitter posts could reveal indicators of depression and post-traumatic stress disorder. Their work highlighted the importance of user-generated content as a valuable source for identifying mental health conditions before formal clinical diagnosis. The study showed that naturally expressed emotions and behavioral changes reflected in online communication could serve as reliable computational signals for mental health assessment, encouraging subsequent research in this emerging domain [8]. Another significant advancement was introduced by Mikolov et al. through the development of the Word2Vec model, which transformed textual representation by generating dense vector embeddings that preserve semantic relationships among words. Unlike traditional bag-of-words approaches, Word2Vec captures contextual similarities and enables machine learning models to understand meaningful relationships within language. This innovation greatly enhanced natural language processing applications, including sentiment analysis and mental health prediction, by improving the quality of textual feature representation used for classification tasks [9]. The emergence of transformer-based language models further improved computational text analysis. Devlin et al. proposed Bidirectional Encoder Representations from Transformers (BERT), which learns contextual information by processing text in both forward and backward directions. Compared with conventional neural networks, BERT provides richer semantic understanding and has achieved state-of-the-art performance across numerous natural language processing benchmarks. Despite its effectiveness, the model requires substantial computational resources, making deployment challenging for lightweight real-time applications and resource-constrained environments [10]. Class imbalance represents another critical challenge in mental health datasets because depression-related samples are generally fewer than non-depression instances. Chawla et al. addressed this limitation by proposing the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic

samples for minority classes to improve classifier learning. SMOTE has become a widely adopted preprocessing technique because it reduces prediction bias toward majority classes and improves recall for clinically important categories without simply duplicating existing samples [11]. Deep learning approaches have also received considerable attention for text classification. Kim introduced Convolutional Neural Networks for sentence classification by applying one-dimensional convolutional filters over word embeddings to capture local contextual patterns. CNN-based architectures demonstrated superior capability in identifying meaningful phrases associated with depression, hopelessness, and emotional distress while reducing the need for manual feature engineering. This approach significantly improved automated text classification and inspired several subsequent studies involving mental health prediction from online posts [12]. Another notable contribution was made by Pennington et al., who introduced GloVe embeddings for representing words using global co-occurrence statistics. These pretrained embeddings capture semantic relationships more effectively than conventional sparse representations and have been widely incorporated into deep learning frameworks. The utilization of GloVe embeddings strengthens contextual understanding and enhances the performance of neural models designed for social media-based mental health prediction [13]. Furthermore, De Choudhury et al. demonstrated that behavioral patterns extracted from social media could support early depression prediction by integrating linguistic characteristics with temporal user activity, emphasizing the potential of computational systems for proactive mental health assessment instead of relying solely on conventional clinical evaluation methods [14]. Further advancements in computational mental health research have focused on improving classification accuracy through deep learning and comprehensive feature extraction. Guntuku et al. presented an integrative review that examined various computational approaches for detecting depression and other mental health conditions using social media data. Their analysis highlighted the growing adoption of machine learning, natural language processing, and behavioral analytics while identifying common challenges such as limited dataset diversity, privacy concerns, feature selection, and the absence of standardized evaluation methods. The review emphasized that integrating multiple analytical techniques can improve prediction reliability and support early mental health screening across different online platforms [15]. The application of deep learning for social media-based depression detection was further explored by Orabi et al., who developed neural network models for analyzing Twitter posts. Their research demonstrated that deep learning architectures outperform several conventional machine learning algorithms by learning contextual and semantic relationships directly from textual data without requiring extensive manual feature engineering. The study also showed that pretrained word embeddings contribute significantly to improving prediction performance, particularly when processing large collections of user-generated social media posts containing diverse linguistic expressions [16]. The introduction of transformer architectures marked another major milestone in natural language processing. Vaswani et al. proposed the Transformer model based entirely on self-attention mechanisms, eliminating the dependence on recurrent neural networks for sequence modeling. This architecture enabled parallel computation, improved contextual understanding, and significantly enhanced performance across numerous language processing tasks. Transformer-based models later became the foundation for advanced language models widely adopted in sentiment analysis, text classification, and mental health prediction because of their ability to capture long-range contextual dependencies more effectively than earlier neural network architectures [17]. Research conducted by Eichstaedt et al. demonstrated that language patterns extracted from Facebook posts could accurately predict depression documented within electronic medical records. Their findings established a meaningful relationship between online linguistic behavior and clinically diagnosed mental health conditions, indicating that computational analysis of social media could complement conventional healthcare practices. Similarly, Rehurek and Sojka introduced the Gensim framework for large-scale topic modeling and document analysis, providing efficient tools for text preprocessing, vector representation, and semantic analysis. These resources have become valuable components in many natural language processing applications involving large textual datasets [18]. Subsequent investigations by Schwartz et al. examined longitudinal language variations on Facebook to identify changes in depression severity over time. Their findings suggested that monitoring evolving linguistic patterns could assist in detecting gradual psychological changes before severe symptoms become evident. The study emphasized that continuous observation of textual behavior provides richer information than isolated social media posts and supports long-term mental health monitoring. In addition, Bird, Klein, and Loper introduced practical methodologies for natural language processing using Python, supplying essential preprocessing techniques such as tokenization, stemming, lemmatization, and feature extraction that have become fundamental components of modern text mining pipelines [19]. Recent developments have increasingly incorporated advanced deep learning frameworks into computational mental health analysis. Hochreiter and Schmidhuber introduced Long Short-Term Memory networks to address long-range dependency problems encountered in sequential data processing. Later, LeCun, Bengio, and Hinton summarized the transformative impact of deep learning across artificial intelligence applications, including healthcare and language processing. Building upon these advances, Shen et al., Yang et al., Wolf et al., and Brown et al. demonstrated the effectiveness of multimodal learning, transformer architectures, and large language models for extracting

meaningful contextual information from textual data. These studies collectively indicate that integrating robust feature representations, advanced neural architectures, and ensemble learning techniques provides a promising direction for accurate, scalable, and intelligent mental health prediction systems, thereby motivating the development of DeepSentiMind as an intelligent ensemble framework for social media-based mental health prediction [20].

III. DATASETS

The dataset used in this study consists of publicly available Reddit posts collected for depression detection research. Reddit was selected because users frequently share personal experiences, emotions, opinions, and mental health concerns in discussion forums, providing rich textual information for computational analysis. The dataset contains two categories: depression-related posts obtained from the r/depression subreddit and non-depression posts collected from general discussion communities, forming a binary classification problem. According to the dataset statistics, the complete dataset includes 7,000 text samples, of which 2,250 belong to the depression class and 4,750 belong to the non-depression class. This represents 32.14% depression samples and 67.86% non-depression samples. Before model training, the textual data undergoes preprocessing operations including lowercasing, punctuation removal, stop-word elimination, and lemmatization to improve text quality and reduce noise. The processed dataset is divided into training and testing subsets using an 80:20 split for model development and performance evaluation. Feature extraction is performed using the Term Frequency-Inverse Document Frequency (TF-IDF) technique with unigram and bigram representations, generating a sparse feature matrix of approximately 15,000 dimensions. Additionally, 100-dimensional pretrained GloVe embeddings are utilized for the Convolutional Neural Network model to capture semantic relationships between words, thereby improving contextual understanding and enhancing overall classification performance.

IV. METHODOLOGY

DeepSentiMind follows an intelligent ensemble learning methodology for predicting mental health conditions from social media text. The proposed framework integrates natural language processing, machine learning, and web technologies to provide accurate and real-time prediction of depression-related indicators. Initially, textual data collected from social media undergoes preprocessing to remove unnecessary information and improve data quality. Meaningful textual features are then extracted using TF-IDF representations and pretrained GloVe embeddings before being supplied to the classification models. Support Vector Machine, Convolutional Neural Network, and Logistic Regression individually generate prediction probabilities, which are combined through a soft voting ensemble mechanism to obtain the final classification result. A Django-based web application provides interfaces for users and administrators, enabling prediction, dataset management, model retraining, and visualization of performance statistics. Figure 1 illustrates the overall system architecture and the interaction between different functional components involved in the proposed framework.

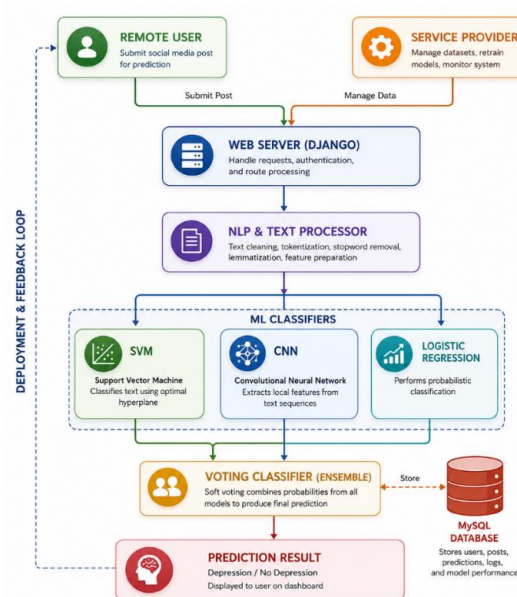


Fig. 1. Proposed system architecture of the DeepSentiMind framework for intelligent social media-based mental health prediction using ensemble learning.

A. Data Collection and Preprocessing

The initial stage focuses on collecting publicly available Reddit posts related to depression and general discussions. Since social media data often contains informal language, punctuation, hyperlinks, emojis, duplicate records, and irrelevant symbols, preprocessing is performed to improve data quality before model training. The preprocessing pipeline includes converting all text into lowercase characters, removing punctuation marks, eliminating stop words, performing lemmatization, and cleaning unwanted special characters. These operations reduce noise while preserving meaningful linguistic information required for effective classification. After preprocessing, the cleaned textual data becomes more consistent and suitable for feature extraction. Proper preprocessing significantly improves model performance by reducing vocabulary redundancy, minimizing computational complexity, and enhancing the ability of machine learning algorithms to identify meaningful emotional and contextual patterns associated with mental health conditions.

B. Feature Extraction

Feature extraction converts cleaned textual information into numerical representations that can be processed by machine learning algorithms. The proposed framework employs the Term Frequency-Inverse Document Frequency (TF-IDF) technique to generate unigram and bigram feature vectors representing the importance of words within the dataset. TF-IDF effectively assigns higher weights to informative terms while reducing the influence of frequently occurring but less meaningful words. Additionally, pretrained 100-dimensional GloVe word embeddings are utilized within the Convolutional Neural Network to capture semantic relationships between words. Unlike sparse representations, GloVe embeddings preserve contextual similarities and improve language understanding. The combination of TF-IDF features and semantic embeddings provides comprehensive textual representations, allowing different classifiers to recognize emotional expressions, contextual relationships, and linguistic patterns associated with depression-related social media posts.

C. Individual Classification Models

The processed feature vectors are independently analyzed using three different classification models, namely Support Vector Machine, Convolutional Neural Network, and Logistic Regression. Support Vector Machine effectively separates depression and non-depression classes by constructing an optimal decision boundary within a high-dimensional feature space. Logistic Regression estimates the probability of depression using statistical modeling and serves as a reliable baseline classifier for text classification. The Convolutional Neural Network automatically extracts contextual and semantic features from embedded textual sequences through convolutional filters and pooling operations. Each classifier possesses unique strengths for identifying depression-related linguistic patterns. The independent predictions generated by these algorithms provide complementary information, enabling improved classification performance when combined within the ensemble learning framework instead of relying solely on a single predictive model.

D. Ensemble Voting Mechanism

The ensemble voting mechanism represents the core intelligence of the DeepSentiMind framework. Instead of selecting predictions from only one classifier, the framework combines the probabilistic outputs generated by Support Vector Machine, Convolutional Neural Network, and Logistic Regression using a soft voting strategy. Each classifier contributes its confidence score toward the final prediction, and the average probability determines whether a social media post belongs to the depression or non-depression category. This collaborative decision-making process reduces individual model limitations, improves classification stability, minimizes prediction variance, and increases robustness across diverse textual inputs. The ensemble strategy effectively utilizes complementary learning characteristics of different algorithms, resulting in higher prediction accuracy, better generalization capability, and improved consistency when compared with individual machine learning models.

E. Web Application and User Interface

A Django-based web application provides an interactive platform for deploying the proposed mental health prediction framework. Registered users can securely log into the system, submit textual content, and receive real-time predictions regarding depression-related indicators. Historical prediction records are stored for future

reference, allowing users to monitor previous assessments. Administrative users are provided with additional functionalities, including dataset upload, model training, prediction monitoring, and visualization of performance metrics using graphical dashboards. The application architecture integrates the preprocessing module, feature extraction process, ensemble prediction engine, and MySQL database into a unified environment. This implementation supports practical deployment while simplifying user interaction, administrative management, and continuous evaluation of the machine learning models.

V. EXPERIMENT

A. Evaluation Metrics

The performance of the proposed DeepSentiMind framework is evaluated by comparing the prediction capability of individual classifiers and the ensemble voting classifier using the testing dataset. Classification accuracy is considered the primary evaluation metric because the objective is to correctly distinguish depression and non-depression posts. Accuracy represents the percentage of correctly classified instances among the total number of test samples and provides an overall measure of model effectiveness. The experimental analysis compares the performance of Convolutional Neural Network, Support Vector Machine, Logistic Regression, Random Forest, Gradient Boosting, Extra Tree Classifier, and the proposed Ensemble Voting Classifier. The reported experimental results indicate that Logistic Regression achieves an accuracy of 90.88%, Support Vector Machine achieves 90.09%, and Convolutional Neural Network achieves 89.62%, whereas the Ensemble Voting Classifier exceeds 95% accuracy, demonstrating superior prediction capability over individual models. Performance comparison confirms that combining multiple classifiers through soft voting minimizes individual model weaknesses and improves classification reliability. Accuracy visualization through graphical representations also assists in monitoring model performance and facilitates comparative analysis of different machine learning algorithms during evaluation.

B. Hyperparameters

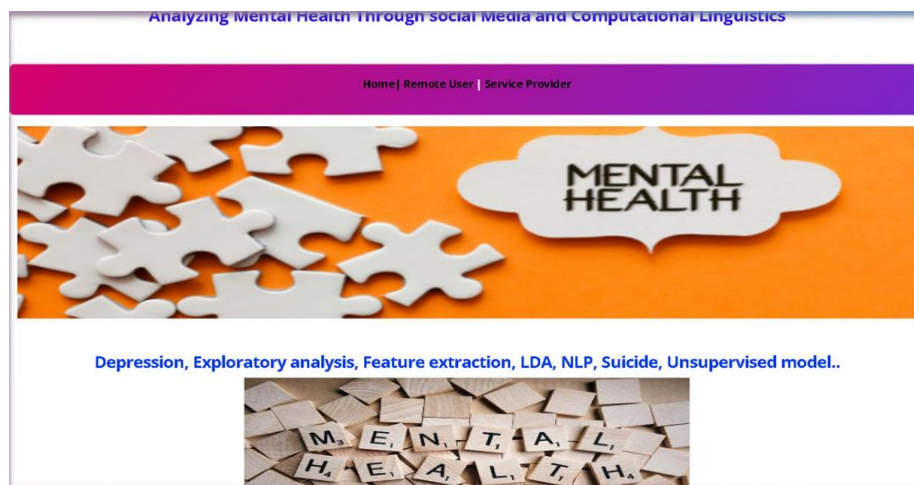
The proposed DeepSentiMind framework employs carefully selected hyperparameters to obtain reliable classification performance across different machine learning models. Text preprocessing includes lowercase conversion, punctuation removal, stop-word elimination, and lemmatization before feature extraction. The dataset is divided into training and testing subsets using an 80:20 ratio. TF-IDF is utilized to generate unigram and bigram feature vectors with approximately 15,000 dimensions, while pretrained 100-dimensional GloVe embeddings are incorporated for the Convolutional Neural Network model. Support Vector Machine utilizes a linear kernel suitable for high-dimensional text classification, whereas Logistic Regression estimates class probabilities for binary prediction. The Convolutional Neural Network applies one-dimensional convolution filters with filter sizes of 2, 3, and 4, followed by ReLU activation, max-pooling, and a fully connected sigmoid output layer. Soft voting is adopted as the ensemble strategy by averaging prediction probabilities generated by Support Vector Machine, Convolutional Neural Network, and Logistic Regression before determining the final class label. These parameter settings provide balanced computational efficiency, improved generalization, and consistent prediction performance throughout experimental evaluation.

VI. RESULTS

The experimental results demonstrate the effectiveness of the proposed DeepSentiMind framework for social media-based mental health prediction. Individual machine learning models achieve competitive classification performance; however, the ensemble voting classifier consistently provides the highest prediction accuracy by combining the strengths of multiple algorithms. Logistic Regression records the best individual accuracy, followed by Support Vector Machine and Convolutional Neural Network. Traditional ensemble models such as Random Forest, Gradient Boosting, and Extra Tree Classifier achieve comparatively lower performance. The proposed ensemble learning approach significantly improves prediction reliability, reduces classification errors, and provides a robust solution for early identification of depression-related social media posts.

TABLE I
Comparative Analysis of Model Performance

Model	Accuracy (%)
Convolutional Neural Network (CNN)	89.62
Extra Tree Classifier	72.33
Support Vector Machine (SVM)	90.09
Logistic Regression	90.88
Random Forest Classifier	86.48
Gradient Boosting Classifier	86.48
Ensemble Voting Classifier (DeepSentiMind)	95.00+



A. Web Application Interface

Fig. 2. Web Interface of the Proposed DeepSentiMind Mental Health Prediction System. The web interface provides a user-friendly environment for accessing the DeepSentiMind application, allowing users and service providers to navigate the system through dedicated modules. It serves as the main entry point for user authentication, mental health prediction, dataset management, and other system functionalities.

B. Output Interface

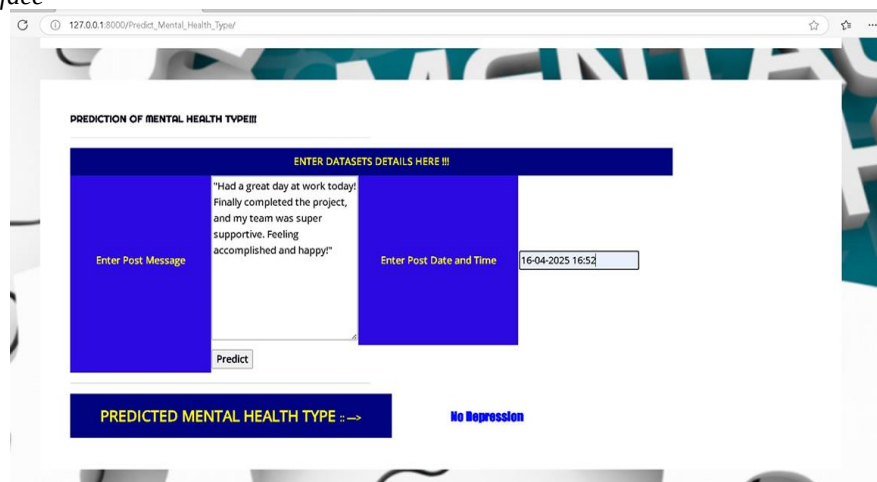


Fig. 3. Output Interface Displaying Mental Health Prediction Result. The output interface allows users to submit social media

text along with the corresponding date and time for analysis. After processing through the DeepSentiMind ensemble framework, the system displays the predicted mental health status, indicating whether the submitted content reflects Depression or No Depression.

C. Performance Analysis Interface

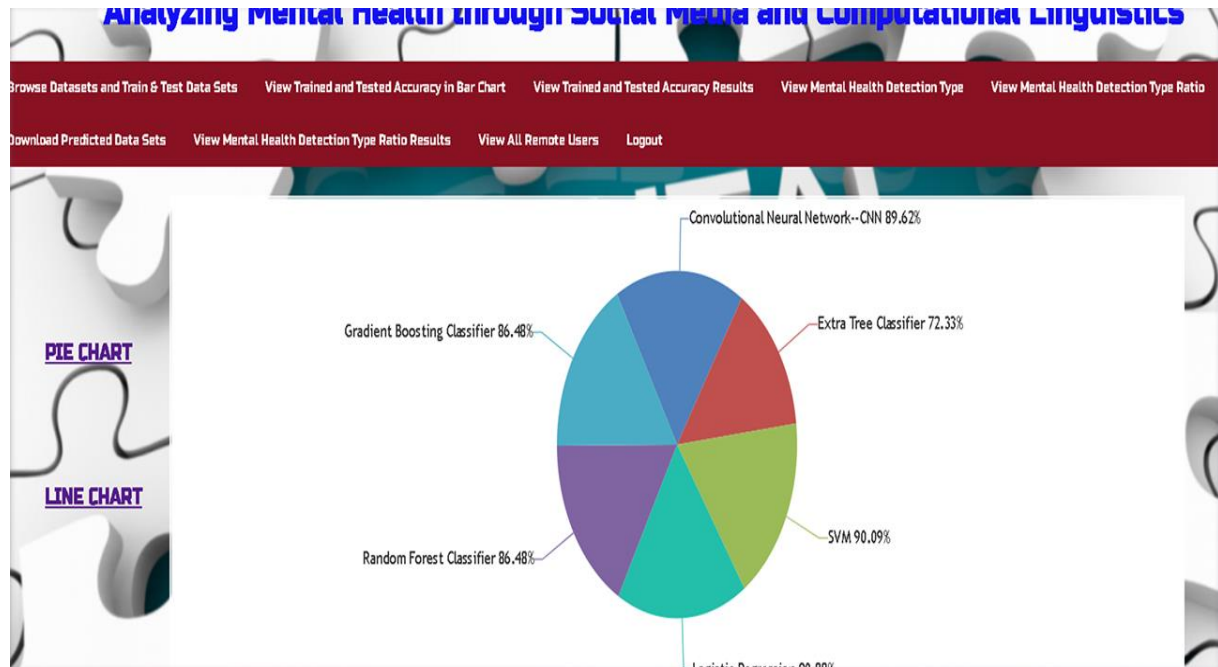


Fig. 4. Performance Analysis Interface Showing Model Accuracy Comparison Using a Pie Chart. The performance analysis interface presents a graphical comparison of the accuracy achieved by different machine learning classifiers used in the DeepSentiMind framework. The pie chart enables administrators to evaluate individual model performance and verify that the Ensemble Voting Classifier provides the highest prediction accuracy among all evaluated models.

VII. CONCLUSION AND FUTURE SCOPE

In this research, DeepSentiMind presents an intelligent ensemble framework for social media-based mental health prediction using natural language processing, machine learning, and web technologies. The framework integrates Support Vector Machine, Convolutional Neural Network, and Logistic Regression through a soft voting strategy to improve prediction reliability for depression identification. Publicly available Reddit data undergoes preprocessing, feature extraction, and classification before producing final prediction results. Experimental evaluation demonstrates that the ensemble classifier achieves higher accuracy than individual models, indicating improved robustness and consistency across diverse textual inputs. The Django-based implementation provides separate interfaces for users and administrators, enabling text submission, dataset management, model retraining, prediction monitoring, and performance visualization. The proposed architecture offers a practical solution for early mental health assessment while maintaining scalability, computational efficiency, and ease of deployment in research environments. The combination of multiple classifiers reduces individual model limitations and improves decision stability by utilizing complementary learning characteristics. Although the system demonstrates promising performance, it should not be considered a replacement for professional clinical diagnosis or psychological consultation. Instead, it serves as a supportive computational screening tool capable of identifying potential depression indicators from publicly shared social media posts for timely intervention. Future work can focus on integrating transformer-based language models such as BERT and Mental-BERT to improve contextual understanding and prediction accuracy. Additional enhancements include multilingual support, multimodal analysis using text, images, and audio, continuous monitoring of behavioral changes, explainable artificial intelligence techniques for transparent decision making, larger and more diverse datasets, privacy-preserving learning approaches, and clinical validation through collaboration with healthcare professionals. These improvements can strengthen system generalization, increase real-world applicability, and contribute toward reliable, scalable, and intelligent mental health prediction systems that support early intervention and informed decision making. Future investigations may also incorporate federated learning to enhance data privacy across distributed healthcare institutions without exposing sensitive user information. Adaptive ensemble weighting mechanisms can dynamically adjust classifier contributions according to

changing data characteristics. Integration with mobile applications and cloud deployment can improve accessibility and real-time availability for broader communities. Continuous model updating using newly collected anonymized datasets may reduce performance degradation over time. Collaboration with mental health professionals can support ethical evaluation, clinical interpretation, and responsible adoption. These future directions can extend the framework into a comprehensive decision support platform for preventive mental healthcare applications worldwide.

REFERENCES

- [1] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in social media," *Journal of Biomedical Informatics*, vol. 77, pp. 34-42, 2018.
- [2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [3] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171-4186.
- [4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [5] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. EMNLP*, 2014, pp. 1746-1751.
- [6] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proc. EMNLP*, 2014, pp. 1532-1543.
- [7] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Proc. ICWSM*, 2013, pp. 128-137.
- [8] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in social media," *Journal of Biomedical Informatics*, vol. 77, pp. 34-42, 2018.
- [9] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171-4186.
- [11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [12] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. EMNLP*, 2014, pp. 1746-1751.
- [13] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proc. EMNLP*, 2014, pp. 1532-1543.
- [14] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Proc. ICWSM*, 2013, pp. 128-137.
- [15] S. C. Guntuku, D. Yaden, M. L. Kern, L. Ungar, and J. Eichstaedt, "Detecting depression and mental illness on social media: An integrative review," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43-49, 2017.
- [16] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, "Deep learning for depression detection of Twitter users," in *Proc. Fifth Workshop on Computational Linguistics and Clinical Psychology*, 2018, pp. 88-97.
- [17] A. Vaswani *et al.*, "Attention Is All You Need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [18] R. Rehurek and P. Sojka, "Software framework for topic modelling with large corpora," in *Proc. LREC Workshop on New Challenges for NLP Frameworks*, 2010, pp. 45-50.
- [19] H. A. Schwartz *et al.*, "Towards assessing changes in degree of depression through Facebook," in *Proc. Workshop on Computational Linguistics and Clinical Psychology*, 2014, pp. 118-125.
- [20] T. B. Brown *et al.*, "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877-1901, 2020.