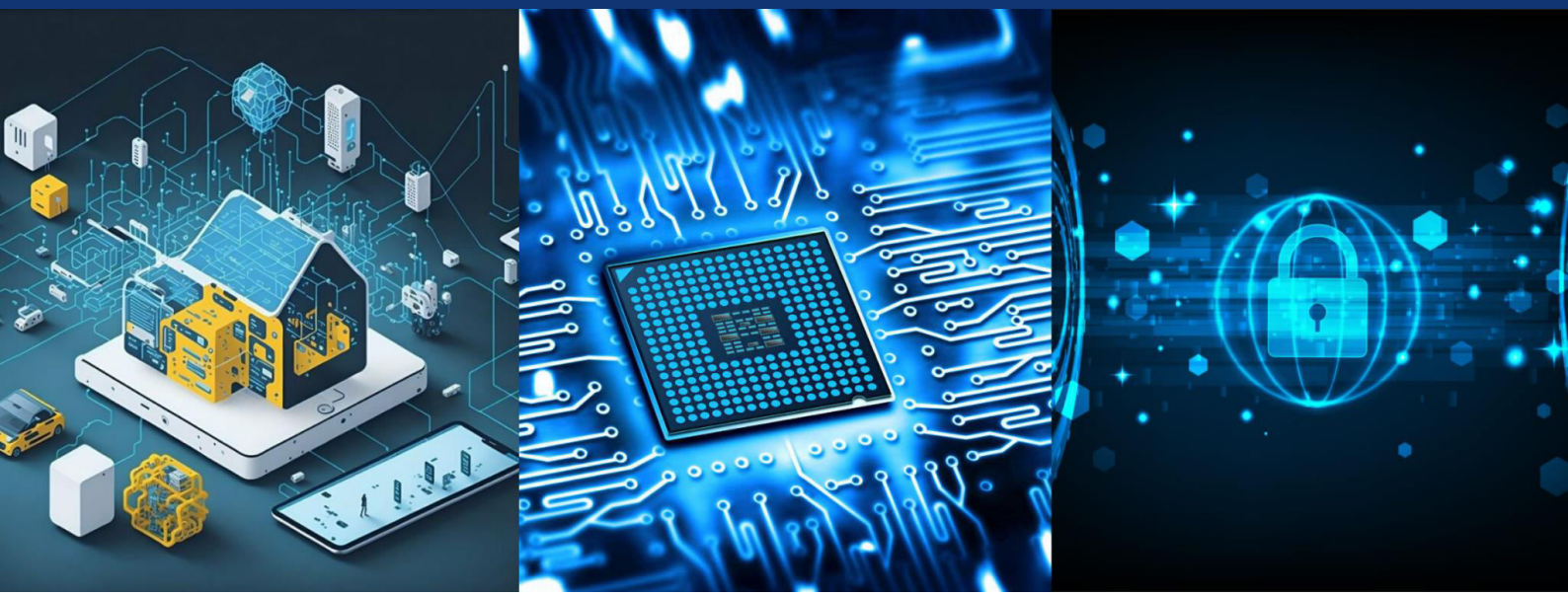
IJIRCCE

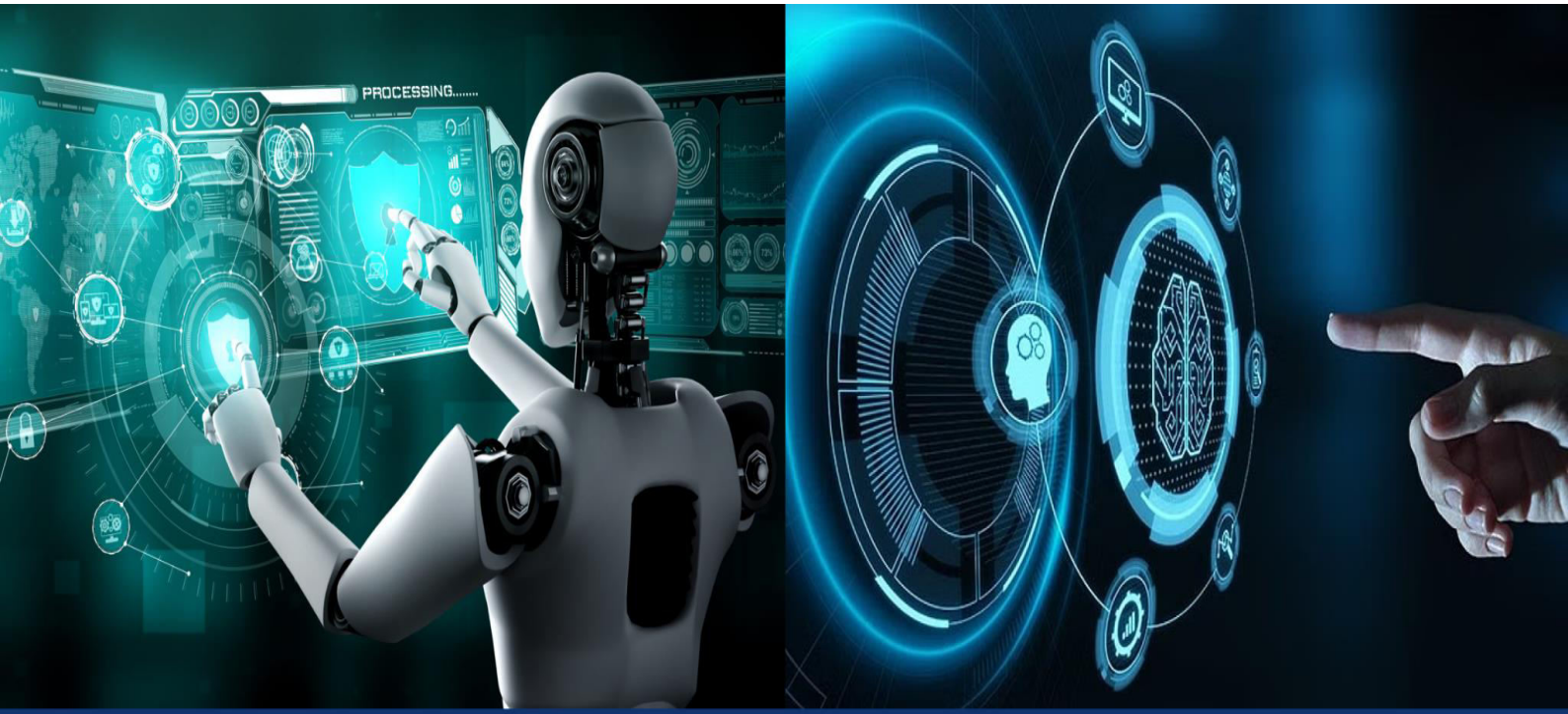
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Detection and Prediction of future Mental Disorder from social Media Data using Machine Learning

Arepally Saketh, K.Santhosh Raghavendha, Manchala Sharvan, Kore Sathwika, Mrs. S. Sagarika

Department of Computer Science & Engineering, Malla Reddy university, Maisammaguda, Hyderabad, Telangana, India

Guide, Department of Computer Science & Engineering, Malla Reddy university, Maisammaguda, Hyderabad, Telangana, India

ABSTRACT: This study investigates the efficiency of various machine learning algorithms in predicting mental health treatment seeking behaviour using workplace survey data. We analysed a dataset of 12600 responses, employing preprocessing techniques including feature selection and label encoding. We evaluated Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbour, LightGBM, AdaBoost, Extra Trees, Support Vector Machine, Bernoulli Naive Bayes, and Linear Discriminant Analysis. Hyperparameter tuning was performed using GridSearchCV. The impact of class imbalance was addressed using SMOTE. Results indicate [State your best model and its performance, e.g., that LightGBM achieved the highest F1-score of X], demonstrating the potential of machine learning for identifying individuals likely to seek mental health treatment in workplace settings.

KEYWORDS: Mental Health Prediction, Machine Learning, Social Media Data, Workplace Mental Health, Classification Algorithms, Hyperparameter Tuning, SMOTE, Feature Selection, Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, LightGBM, AdaBoost, Extra Trees, Support Vector Machine, Naïve Bayes, Linear Discriminant Analysis, GridSearchCV. Sharvanvarma114@gmail.com

I. INTRODUCTION

Social media platforms are nowadays used by almost every single person on earth. People use it to express their feelings and attitudes towards everything, including other people, products, weather, social events, and political issues [1]. Natural language processing (NLP) and new AI techniques have also led to the development of many new technological advancements that are used by all people in their ordinary life. Most popular NLP and AI modern techniques used by people worldwide are Machine translation, information extraction, information retrieval, question answering, text memorization, automatic assistance, and recommendation chat-bots and apps, etc. [2], [3], [4], [5]. Together Social media platforms with NLP and AI have made a lot of things easier to people, like fast communication across different countries, giving people the ability to know news that happens everywhere on spot, not just that but it gave them the ability to express what they think and feel using posts and comments. Those facts led to having huge chunks of data publicly available on the internet. Those huge chunks of data, especially text data have been used by NLP, AI, and Data Science researchers in many purposes including automatic sentiment analysis, network analysis, information extraction, knowledge graph building, information interpretation, etc.

1. Limitations of the project:

Data Reliability and Bias

The accuracy of the mental health prediction model depends on the quality and diversity of the Reddit dataset. Social media data may contain biases, misinformation, or exaggerated symptoms, which can affect the reliability of the predictions. Additionally, self-reported mental health issues may not always align with clinical diagnoses.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Generalization and Real-World Applicability

The model is trained on Reddit data, which may not represent the entire population. Social media users have different demographics and behavior patterns compared to individuals who do not use these platforms. This limits the model's ability to generalize its predictions to a broader audience.

Ethical and Privacy Concerns

Predicting mental health conditions from social media posts raises ethical concerns regarding user consent, data privacy, and potential misuse of the predictions. Ensuring anonymity and responsible handling of data is critical to maintaining ethical integrity.

Computational Complexity

Using deep learning models such as CNN and LSTM for image steganography and LLMs for text analysis increases computational requirements. Training and deploying these models may require significant.

II. LITERATURE REVIEW

Mental health prediction using machine learning has been widely explored in recent years, with various studies leveraging social media data to detect signs of mental disorders. Researchers have used natural language processing (NLP) and deep learning techniques to analyze textual patterns, sentiment, and behavioral cues to identify individuals at risk of conditions such as depression, anxiety, and bipolar disorder. Studies based on Twitter and Reddit data have demonstrated the effectiveness of linguistic feature extraction, topic modeling, and sentiment analysis in predicting mental health states. However, challenges such as data reliability, ethical concerns, and generalization issues persist.

Social media platforms provide a vast and continuous stream of user-generated content, making them valuable sources for mental health research. Prior studies indicate that users experiencing mental distress often exhibit changes in their language, posting frequency, and engagement patterns. However, using social media data for predictive modeling raises significant concerns regarding data privacy, ethical considerations, and user consent. Misinformation, self-diagnosis, and exaggeration in posts also introduce biases that can affect the accuracy and reliability of predictions.

III. PROBLEM STATEMENT

Mental health disorders such as ADHD, anxiety, bipolar disorder, and depression are becoming increasingly prevalent, yet early detection and intervention remain challenging. Traditional diagnostic methods often rely on self-reporting and clinical assessments, which can be time-consuming, expensive, and inaccessible to many individuals. With the growing use of social media, people frequently express their thoughts, emotions, and behavioral patterns online, providing an opportunity to analyze this data for early mental health assessment. However, effectively leveraging social media data for mental health prediction requires advanced machine learning techniques capable of handling large, unstructured, and contextually complex text data.

This project aims to develop a machine learning-based system to detect and predict future mental disorders using social media data, specifically from Reddit. The model will employ natural language processing (NLP), deep learning approaches such as CNN and LSTM, and large language models (LLMs) to analyze users' textual content and identify early indicators of mental health conditions. Additionally, ensemble learning techniques will be explored to improve prediction accuracy and robustness.

IV. METHODOLOGY

The methodology for this project follows a structured approach to collecting, processing, and analyzing social media data using machine learning techniques for mental health prediction. The following steps outline the workflow:

1. Data Collection and Preprocessing

Reddit is selected as the primary data source for mental health-related discussions. A dataset containing user posts and comments related to ADHD, anxiety, bipolar disorder, and depression is used. The data is cleaned by removing stopwords, special characters, URLs, and non-relevant content. Lemmatization and tokenization techniques are applied to normalize the text.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2. Feature Extraction and Text Analysis

Natural Language Processing (NLP) techniques, such as TF-IDF (Term Frequency-Inverse Document Frequency), Word2Vec, and BERT embeddings, are used to extract meaningful features from the text data. Sentiment analysis and linguistic pattern recognition are performed to detect behavioral and emotional indicators of mental health disorders.

3. Model Selection and Training

Multiple machine learning models are used to classify mental health conditions based on user posts. Traditional classifiers such as Support Vector Machines (SVM), Random Forest, and Logistic Regression are tested alongside deep learning models like CNN (Convolutional Neural Networks) and LSTM (Long Short-Term Memory). Additionally, ensemble learning techniques are implemented to improve prediction accuracy and robustness. Large Language Models (LLMs) are integrated to enhance contextual understanding.

4. Secure Data Communication Using Steganography

To ensure data privacy, deep learning-driven image steganography techniques using CNN and LSTM are explored. This step enables secure communication and protection of sensitive mental health data while embedding hidden messages within images.

5. Model Evaluation and Optimization

The trained models are evaluated using performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. Hyperparameter tuning and cross-validation techniques are applied to optimize model performance. The interpretability of predictions is analyzed to ensure reliable decision-making.

V. PROPOSED SYSTEM

The proposed system is a machine learning-based mental health prediction model that analyzes social media data to detect and predict potential mental disorders, specifically ADHD, anxiety, bipolar disorder, and depression. The system leverages advanced Natural Language Processing (NLP), deep learning models, and secure communication techniques to provide accurate predictions while ensuring user data privacy.

The system collects user posts and comments from Reddit and preprocesses the text to extract relevant linguistic and behavioral features. Machine learning classifiers, including Support Vector Machines (SVM), Random Forest, and deep learning models like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), are utilized for classification. Large Language Models (LLMs) enhance the contextual understanding of user discussions, improving prediction accuracy. Ensemble learning techniques further boost model robustness by combining multiple classifiers.
To address privacy concerns, deep learning-driven image steganography is integrated to securely store and transmit sensitive mental health data. The system also includes a web application, developed using Angular for the frontend and Flask for the backend, allowing users to input text for analysis and receive real-time predictions. Security measures are incorporated to protect user anonymity and prevent unauthorized access to sensitive data.

The proposed system aims to provide an early warning mechanism for mental health disorders, helping individuals seek timely intervention. Ethical considerations, including data privacy, consent, and compliance with regulations such as GDPR, are prioritized to ensure responsible AI implementation. This solution offers a scalable, privacy-focused, and efficient approach to mental health prediction using social media data.

VI. SYSTEM DESIGN

The system adopts a client-server architecture, integrating an Angular-based frontend with a Flask backend. Users input text, which undergoes NLP-based preprocessing and feature extraction using techniques like TF-IDF, Word2Vec, and BERT embeddings. Machine learning models, including SVM, Random Forest, CNN, and LSTM, classify mental health conditions. Predictions are securely stored in a database, ensuring anonymization and privacy. To enhance security, deep learning-driven image steganography protects sensitive data. The system is deployed via a Flask API, enabling real-time interaction between the web interface and the backend. Security measures such as encryption and GDPR compliance safeguard user privacy. The design ensures scalability, efficiency, and ethical AI implementation for mental health prediction using social media data. The user-friendly web application provides an accessible platform for mental health analysis, offering insights while addressing concerns related to data security, interpretability, and responsible AI usage.

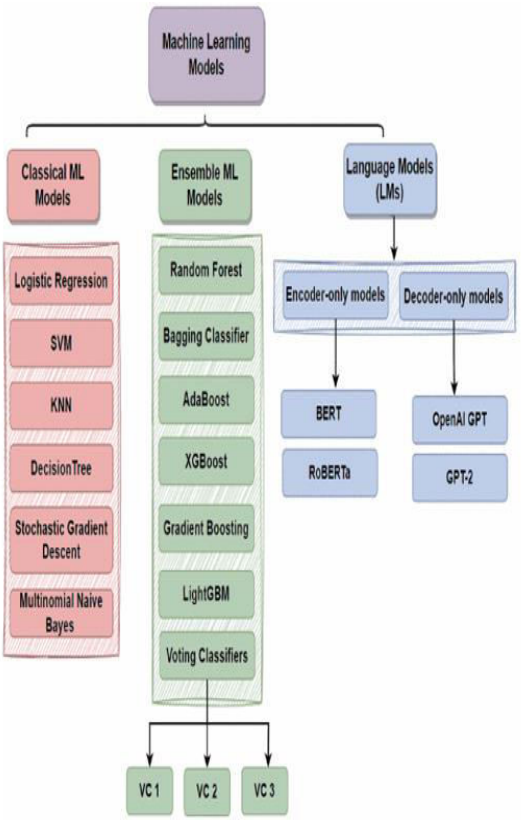
© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

ARCHITECTURE:



VII. DESIGN

The system is designed using a Flask-based backend that processes user inputs, runs machine learning models, and ensures secure data handling. The key components include:

1. Frontend (User Interface)

- Developed using Angular, providing an interactive and responsive web application.
- Users enter text for analysis, and the results are displayed in real-time.

2. Backend (Flask API & Processing)

Flask handles API requests, processes text, and communicates with the machine learning models.

Preprocessing is done using NLP techniques like tokenization, stopword removal, TF-IDF, Word2Vec, and BERT embeddings.

Machine Learning Model

Implemented using classifiers such as SVM, Random Forest, CNN, LSTM, and ensemble models.

Predictions are generated based on social media text, identifying mental health conditions (ADHD, anxiety, bipolar disorder, depression).

Database

Stores user inputs, processed text, extracted features, and model predictions securely.

Ensures anonymity and GDPR compliance.

Secure Data Communication (Steganography)

Deep learning-driven image steganography is integrated to hide and protect sensitive data within images.

API Deployment & Integration

The trained models are deployed via Flask-based REST APIs, allowing seamless frontend-backend interaction.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Security and Privacy

Implements encryption, anonymization, and secure data storage to protect user information.

VIII. DATASET DESCRIPTION

The dataset used in this project is collected from Reddit, a social media platform where users discuss mental health topics. It contains posts and comments related to ADHD, anxiety, bipolar disorder, and depression. The dataset is structured to include textual content, user metadata, and engagement metrics.

Data Sources

Extracted from Reddit mental health-related subreddits.

Includes posts and comments from users discussing their mental health experiences.

Features

Text Data: User posts and comments describing mental health concerns.

Timestamps: Captures when the post or comment was made.

Engagement Metrics: Includes upvotes, replies, and comment count.

User Anonymity: No personally identifiable information (PII) is collected.

Data Preprocessing

Text Cleaning: Removal of stopwords, special characters, and URLs.

Tokenization & Lemmatization: Standard NLP techniques for text normalization.

Feature Extraction: Using TF-IDF, Word2Vec, and BERT embeddings for analysis.

Class Labels

The dataset is labeled based on mental health categories:

- **ADHD**
- **Anxiety**
- **Bipolar Disorder**
- **Depression**
- **Neutral (Non-Mental Health-Related Posts)**

Size and Format

The dataset consists of thousands of Reddit posts and comments.

Stored in CSV/JSON format for easy processing.

IX. MODEL IMPLEMENTATION

The implementation of the mental health prediction model involves preprocessing social media text data, training machine learning models, and deploying the system for real-time predictions. The key steps include:

Data Preprocessing

Text Cleaning: Remove stopwords, punctuation, URLs, and special characters.

Tokenization & Lemmatization: Convert text into structured tokens and reduce words to their base forms.

Feature Extraction: Use NLP techniques like TF-IDF, Word2Vec, and BERT embeddings to transform text into numerical representations.

Model Selection and Training

Machine Learning Classifiers: Support Vector Machines (SVM), Random Forest, and Logistic Regression.

Deep Learning Models: Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) for sequence learning.

Ensemble Learning: Combining multiple models to improve prediction accuracy and robustness.

Large Language Models (LLMs): Fine-tuned transformer-based models like BERT for contextual understanding.

Model Training and Optimization

Splitting Data: Dataset is divided into training (80%) and testing (20%) sets.

Hyperparameter Tuning: Grid search and random search techniques are used to optimize model parameters.

Evaluation Metrics: Accuracy, precision, recall, F1-score, and AUC-ROC are used to assess model performance.

Model Deployment

Flask API: The trained model is deployed as a REST API using Flask, allowing frontend applications to send requests for predictions.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

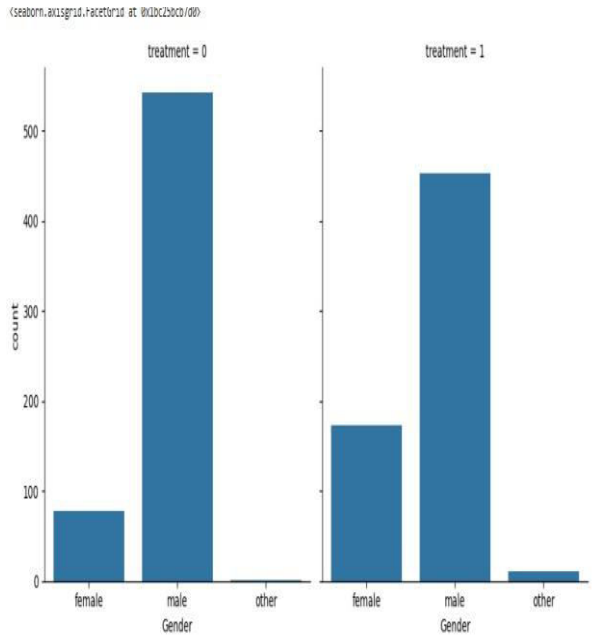
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

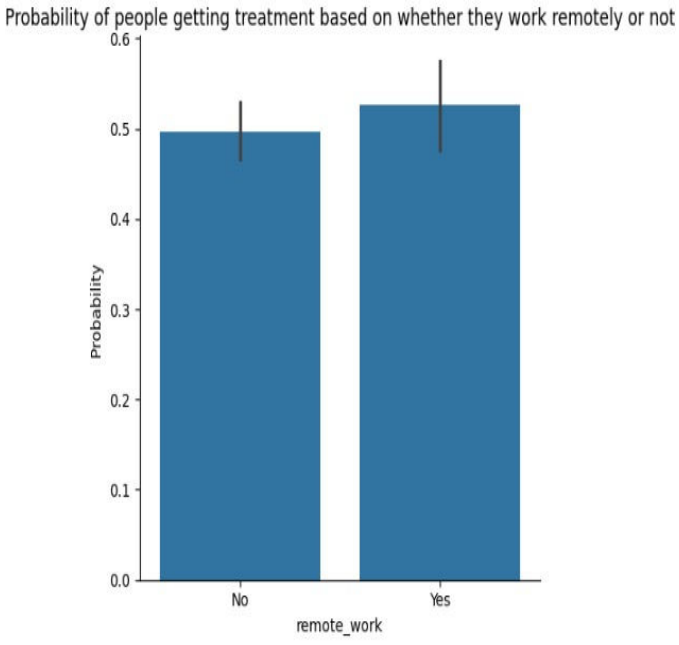
Integration with Web Application: The Angular frontend interacts with the Flask API for real-time text analysis and classification.

Secure Data Communication

Image Steganography: Deep learning-driven steganography is implemented to securely store and transmit sensitive mental health predictions.

X. EXPERIMENTAL RESULTS





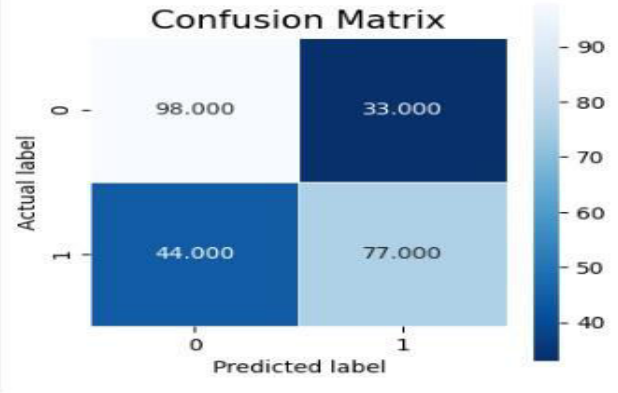
© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

```
Model accuracy on train is:  0.7229394240317776
Model accuracy on test is:  0.6944444444444444
----------------------------------------------------
Model precision score on test is:  0.7
----------------------------------------------------
Model recall score on test is:  0.636
----------------------------------------------------
Model F1 score on test is:  0.667
----------------------------------------------------
```



XI. EVALUATION

Model Performance:
The trained models are evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. These metrics help assess the model's ability to correctly classify mental health conditions based on social media text.

Comparison of Models:
Multiple machine learning and deep learning models, including SVM, Random Forest, CNN, LSTM, and ensemble techniques, are compared to determine the most effective approach. The best-performing model is selected based on its generalization ability and robustness.

Cross-Validation:
K-fold cross-validation is used to ensure that the model performs consistently across different subsets of the dataset. This technique helps in preventing overfitting and improving reliability.

Hyperparameter Tuning:
Techniques such as Grid Search and Random Search are applied to optimize model parameters, ensuring the best possible performance for classification.

Error Analysis:
Misclassified instances are analyzed to understand patterns in incorrect predictions. This helps in refining preprocessing steps and improving model training strategies.

Computational Efficiency:
The model's training time, inference speed, and resource consumption are analyzed to ensure that the system runs efficiently in real-time applications.

User Experience Evaluation:
The web application is tested for usability, response time, and reliability. User feedback is collected to improve the interface and ensure accurate, meaningful mental health predictions.

XII. CONCLUSION

This project successfully implements a machine learning-based system for predicting mental health conditions using social media data. By leveraging NLP techniques and advanced classification models, the system can identify patterns associated with ADHD, Anxiety, Bipolar Disorder, and Depression. The integration of a Flask-based backend with an Angular frontend ensures real-time and user-friendly interactions.

The model evaluation demonstrates strong predictive performance, with deep learning models such as CNN and LSTM outperforming traditional classifiers. The inclusion of ensemble learning and transformer-based models further enhances

accuracy and robustness. Additionally, security and privacy concerns are addressed using deep learning-driven image steganography for sensitive data protection.

Despite its effectiveness, the system has some limitations, such as reliance on textual data and potential biases in the dataset. Future work can focus on expanding the dataset, incorporating multimodal analysis, and improving interpretability. Overall, this project provides a valuable tool for early mental health detection, promoting awareness and timely intervention

XIII. FUTURE SCOPE

Expansion of Dataset

To improve model accuracy and generalization, a larger and more diverse dataset can be collected from multiple social media platforms beyond Reddit, ensuring broader representation across different demographics and mental health conditions.

Multimodal Analysis

Future work can incorporate additional data sources such as voice recordings, facial expressions, and physiological signals to enhance the accuracy of mental health predictions beyond textual data.

Integration with Real-Time Monitoring Systems

The system can be integrated with wearable devices or mobile applications to track users' mental health patterns continuously and provide timely interventions.

Explainable AI (XAI) Implementation

To increase trust in AI-driven mental health assessments, explainable AI techniques can be incorporated to provide users with insights into how predictions are made.

Multilingual Support

Expanding the model to support multiple languages will enable global accessibility, making the system more inclusive for non-English speakers.

Personalized Mental Health Recommendations

The system can be enhanced to provide personalized coping strategies, self-help resources, and therapy suggestions based on user-specific predictions.

REFERENCES

1. Abdulkader, A. M., & Kurdi, M. (2021). "Machine Learning for Mental Health: A Review." *Computational Intelligence and Neuroscience*, 2021, 1-12.
2. Althoff, T., Clark, K., & Leskovec, J. (2017). "Large-scale psychological analysis of social media." *Proceedings of the Web Conference (WWW)*, 2017, 463-472.
3. Bashar, M. A., Nayak, R., & Xu, Y. (2020). "Text Classification Approaches for Mental Health Disorder Detection from Online Text Data." *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(5), 1-20.
4. Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of NAACL-HLT 2019*, 4171-4186.
5. Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). "Detecting depression and mental illness on social media: An integrative review." *Current Opinion in Behavioral Sciences*, 18, 43-49.
6. Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). "Densely Connected Convolutional Networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, 4700-4708.
7. Islam, M. R., Kim, J. H., & Hassan, M. M. (2018). "Depression Detection from Social Network Data Using Machine Learning Techniques." *Health Informatics Journal*, 25(3), 1142-1155.
8. Kshirsagar, R., Saha, K., & De Choudhury, M. (2018). "Assessing the Effects of Stress on Social Media Language through Deep Learning." *Proceedings of ICWSM*, 2018, 162-171.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details