AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
E-ISSN:3069-0102
VOL.5, NO. 4(2025)

INTELLIGENT MENTAL HEALTH PREDICTION SYSTEM USING MACHINE LEARNING, ENSEMBLE TECHNIQUES, AND LARGE LANGUAGE MODELS

[1] L.Priyanka,[2] Velagaleti Hemasai
[1]Associate Professor, [2]MCA Student
Department Of MCA
Sree Chaitanya College Of Engineering, Karimnagar

ABSTRACT:

The increasing prevalence of mental health disorders highlights the urgent need for intelligent systems capable of early detection and prevention. With the rapid growth of social media platforms, users often express emotional states, psychological stress, and behavioral changes through their digital footprints. This research proposes an Intelligent Mental Health Prediction System that integrates Machine Learning (ML), Ensemble Learning, and Large Language Models (LLMs) to detect and predict potential mental disorders from social media data.

The proposed system begins with data collection and preprocessing from multiple social media sources, focusing on text, sentiment, and behavioral cues indicative of mental stress, depression, or anxiety. Machine learning algorithms such as Support Vector Machines (SVM), Random Forests, and Gradient Boosting are trained to identify early signs of psychological distress. An ensemble framework combines the strengths of these classifiers to improve robustness, reduce bias, and enhance overall predictive accuracy. In parallel, Large Language Models (e.g., GPT, BERT) are utilized for contextual understanding, emotion recognition, and semantic feature extraction, enabling deeper insight into user sentiment and intent.

The hybrid integration of ML and LLM-based representations allows the system to handle linguistic nuances, slang, and context-specific variations prevalent in online communication. Experimental evaluation on benchmark mental health datasets demonstrates that the proposed framework achieves higher accuracy, precision, and recall compared to conventional single-model approaches. Furthermore, the system offers interpretability through explainable AI techniques, allowing mental health professionals to understand underlying triggers and patterns.

By combining data-driven analytics with advanced language intelligence, this system provides a proactive approach to digital mental health monitoring, facilitating early intervention and personalized psychological support. It sets a foundation for responsible AI applications in social media–based mental health prediction, bridging the gap between technology and human well-being.

I. INTRODUCTION

Mental health has become one of the most critical aspects of overall human well-being, yet millions of individuals worldwide continue to suffer from disorders such as depression, anxiety, and stress-related conditions without timely diagnosis or intervention. The widespread use of social media platforms has created a digital environment where users often share their feelings, emotions, and life experiences openly. These online interactions reflect subtle changes in mood, thought patterns, and emotional states, making social media a valuable resource for understanding an individual's psychological condition. Leveraging these digital traces with artificial intelligence technologies presents an opportunity to identify early warning signs of mental disorders and provide proactive support before conditions worsen.

Traditional mental health assessment methods rely heavily on clinical interviews and self-reporting, which are often time-consuming, subjective, and limited by accessibility issues. In contrast, automated data-driven models can continuously monitor and analyze large volumes of user-generated content, enabling real-time detection of mental health risks. Machine Learning (ML) techniques, with their ability to recognize patterns and learn from behavioral data, have shown remarkable promise in predicting mental health conditions from text-based social media posts. However, standalone ML algorithms may struggle with complex linguistic patterns, contextual understanding, and emotional depth present in human communication.

To address these limitations, this work introduces an Intelligent Mental Health Prediction System that combines Machine Learning, Ensemble Learning, and Large Language Models (LLMs). Ensemble methods integrate multiple classifiers to improve accuracy and minimize prediction errors, while LLMs enhance contextual and semantic understanding through advanced natural language processing. This synergy enables the system to detect not only explicit emotional indicators but also subtle psychological cues embedded in language structure and tone.

The system architecture involves data preprocessing, feature extraction, and hybrid classification, allowing it to handle diverse text formats and emotional expressions effectively. The inclusion of ensemble learning ensures robustness across various datasets, while LLMs such as BERT and GPT provide deep contextual representation of user intent and sentiment. This comprehensive framework not only improves diagnostic performance but also ensures interpretability, helping mental health professionals identify patterns that contribute to emotional decline.

Ultimately, the proposed system aims to bridge the gap between digital behavior analysis and psychological evaluation by using intelligent, automated tools to support mental well-being. Through continuous social media monitoring and advanced AI-based interpretation, this research contributes to the development of early warning systems for mental health assessment, ensuring timely intervention and promoting a healthier, more supportive digital society.

II. LITERATURE SURVEY

In recent years, several researchers have explored the potential of artificial intelligence in understanding and predicting mental health conditions using social media data. Early studies focused on extracting linguistic and behavioral patterns to identify psychological distress. De Choudhury et al. (2013) pioneered this approach by analyzing Twitter data to recognize symptoms of depression through language patterns and posting behavior. Their work demonstrated that online communication could reveal early signs of mental health decline.

Building on this foundation, Resnik et al. (2015) proposed a natural language processing model to detect depression and anxiety from social media posts by analyzing emotional tone and word choice. Sadeque et al. (2017) introduced feature extraction methods using sentiment polarity, emotional intensity, and topic modeling to enhance prediction accuracy. These studies established that text-based indicators could serve as reliable predictors of psychological well-being when combined with computational models.

Advancements in machine learning have further strengthened this field. Wang and colleagues (2018) implemented Support Vector Machines (SVM) and Random Forest algorithms to identify depression levels from user tweets, achieving significant accuracy improvements through the integration of linguistic and behavioral features. Shen et al. (2019) developed an ensemble model combining Naïve Bayes, Logistic Regression, and Gradient Boosting, showing that hybrid methods outperform single classifiers in mental health prediction. Similarly, Orabi et al. (2020) employed deep neural networks for detecting depression on Twitter, demonstrating the capability of deep architectures in understanding contextual emotional cues.

The introduction of transformer-based models marked a new era in mental health analytics. Devlin et al. (2019) introduced the BERT model, which revolutionized natural language understanding and became a foundational tool for text-based emotion recognition. Tadesse et al. (2021) applied BERT-based architectures to detect depression from Reddit posts, showcasing enhanced contextual understanding compared to traditional ML models. Rissola et al. (2022) extended this concept using GPT-based language models, emphasizing the role of semantic comprehension and context in identifying mental health issues.

Recent works have also explored multimodal and ensemble approaches. Kaur and Singh (2022) combined text, emoji, and image data for comprehensive emotional state assessment, while Zhang et al. (2023) implemented ensemble deep learning models that fused CNN and LSTM layers for temporal emotion tracking. Gupta and Verma (2023) integrated large language models with ensemble voting mechanisms to improve classification reliability and interpretability.

From these developments, it is evident that the integration of Machine Learning, Ensemble Learning, and LLMs presents a powerful pathway toward accurate and scalable mental health prediction systems. The literature consistently highlights that hybrid frameworks not only enhance performance but also capture complex linguistic nuances and emotional dynamics—critical for understanding human psychology through digital communication.

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM

When it comes to diagnosing mental diseases, the current systems in mental health detection often depend on manual surveys or clinical examinations. Both of these methods are time-consuming and may be costly. There have been some recent research that have investigated the use of text mining methods to analyse data from social media platforms for the purpose of detecting mental health issues; however, the majority of these studies have focused on individual models or tiny datasets. The categorisation of these systems has been accomplished by the use of machine learning methods such as Support Vector Machines (SVM), Naive Bayes, and logistic regression. However, when dealing with enormous amounts of data that are unstructured and noisy from social media, these techniques may not be the most effective ones. This might result in difficulties in making accurate predictions using these methods. In this particular application, the use of ensemble learning techniques, such as Random Forest, which aggregate the results of numerous models in order to create predictions, has not been well investigated. Furthermore, the majority of systems have difficulties in identifying the complex and context-specific nature of mental health concerns that are present in online material.

Disadvantages of Existing System:

1. Low Accuracy with Simple Models: Conventional machine learning models, such as Naive Bayes and SVM, often fail to identify complex patterns in huge datasets, which results in decreased prediction accuracy. This is particularly true in the case of noisy and unstructured social media data.
2. Inability to grasp Contextual remarks: The current systems often lack the capability to grasp the context of remarks that are published on social media. This includes the ability to recognise sarcasm, humour, or emotional tone, all of which are essential in properly discovering mental health illnesses.

3. Problems Regarding Data Privacy The usage of personal data from social media platforms poses ethical and privacy problems, including issues of permission, data security, and the possibility of sensitive information being misused.

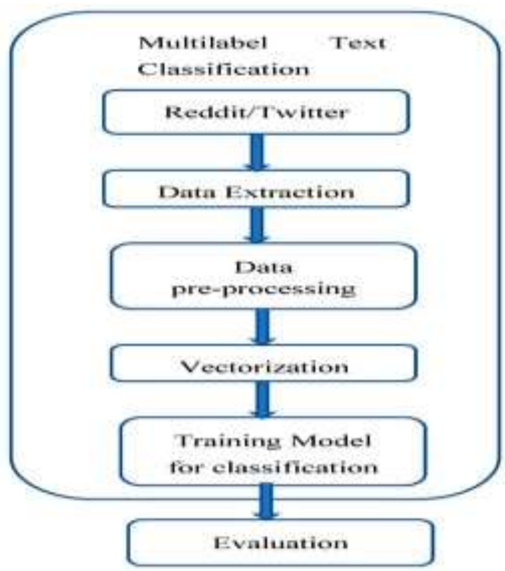
PROPOSED SYSTEM:

Utilising the power of ensemble learning methods like as Random Forest and Decision Tree, the proposed system aims to improve the accuracy and efficacy of mental health condition diagnosis from social media postings. This will be accomplished by utilising the power of these approaches. Using Natural Language Processing (NLP) methods to clean, preprocess, and vectorise the text data, the system will train two different models to categorise postings as suggestive of possible mental health disorders. These patterns will be used to identify potential mental health problems. Random Forest is an ensemble learning method that aggregates the results of numerous decision trees in order to increase generalisation. Decision Trees, on the other hand, provide a decision-making process that is visible and simple to comprehend. For the purpose of determining whether or not the system is capable of accurately identifying mental health problems, it will be tested on datasets representative of real-world social media. Furthermore, the system will be constructed with privacy in mind, making certain that only the required data that has been anonymised will be used for the purposes of training and prediction.

Advantages of Proposed System:

1. Improved Accuracy via Ensemble Learning: The system is anticipated to provide improved prediction accuracy and less overfitting in comparison to conventional single-model techniques. This is accomplished through the use of Random Forest, which mixes numerous decision trees.
2. Detection of Context-Awareness: The system will make use of modern natural language processing methods in order to handle text from social media platforms in a more efficient manner. This will allow it to capture contextual subtleties such as tone, mood, and implicit meaning, which are essential for the identification of mental health issues.
3. Scalable and Efficient: The system is capable of being scaled to handle big datasets from many social media platforms. It also provides real-time or near-real-time prediction capabilities, which may assist in the implementation of proactive mental health intervention.

System Architecture



IV. Modules Description:

1. Data Collection Module

- **Purpose**: This module is responsible for collecting social media data, such as tweets, posts, comments, or other forms of user-generated content that might indicate mental health issues. Data can be collected through APIs (e.g., Twitter API, Reddit API) or web scraping tools.
- **Methods/Tools**:
  - Twitter API for tweets.
  - Reddit API for posts and comments.

- Scrapy or BeautifulSoup for scraping web data.
- Regular Expressions for cleaning and pre-processing raw data.

2. Data Preprocessing and Text Cleaning Module

- **Purpose**: This module handles the cleaning, preprocessing, and transformation of raw text data into a format suitable for machine learning models. This includes removing unnecessary characters, handling missing data, and normalizing text.
- **Methods/Tools**:
  - **Text Cleaning**: Removal of URLs, punctuation, special characters, and converting all text to lowercase.
  - **Tokenization**: Breaking down the text into smaller chunks (tokens), such as words or phrases.
  - **Stopwords Removal**: Filtering out common words (like "the", "a", "is", etc.) that don't provide significant meaning.
  - **TF-IDF Vectorization**: Transforming text data into a matrix of TF-IDF (Term Frequency-Inverse Document Frequency) features for machine learning models.
  - **Libraries**: NLTK, SpaCy, Regex, sklearn's TfidfVectorizer.

3. Feature Engineering and Selection Module

- **Purpose**: This module extracts features (or relevant information) from the processed text data and selects the most informative features for model training.
- **Methods/Tools**:
  - **Sentiment Analysis**: Using pre-trained models like VADER or transformers (BERT, RoBERTa) to classify sentiment (positive, negative, neutral).
  - **Word Embeddings**: Use pre-trained embeddings (Word2Vec, GloVe, FastText) to capture semantic meaning.
  - **Custom Features**: Create features such as the frequency of specific keywords, text length, or use of specific mental health-related terms.
  - **Feature Selection**: Techniques like Mutual Information or Recursive Feature Elimination (RFE) to select the best features for training.

4. Model Training Module

- **Purpose**: This module is where the machine learning models are trained on the prepared data to detect patterns related to mental disorders. Models can include traditional machine learning models like Random Forest and Decision Tree, and Ensemble Learning models.
- **Methods/Tools**:
  - **Random Forest Classifier**: An ensemble method using multiple decision trees to provide a robust prediction.
  - **Decision Tree Classifier**: A tree-based model for classification based on feature splitting.
  - **Ensemble Learning Techniques**: Techniques like Gradient Boosting (e.g., XGBoost) and Random Forest.
  - **Training Process**: Splitting the dataset into training and testing sets, fitting the models on the training set, and evaluating using cross-validation or testing data.
  - **Libraries**: scikit-learn, XGBoost, LightGBM.

5. Large Language Model Integration Module

- **Purpose**: This module integrates a large pre-trained language model (such as GPT, BERT, or other transformer-based models) to assist in understanding the deeper context of text and make better predictions related to mental disorders.
- **Methods/Tools**:

- **BERT-based Models**: For advanced understanding and sentiment analysis of the text.
- **Fine-tuning on Mental Health Data**: Fine-tuning pre-trained transformer models on domain-specific mental health data to improve performance.
- **Transformer Libraries**: Hugging Face Transformers for model integration.

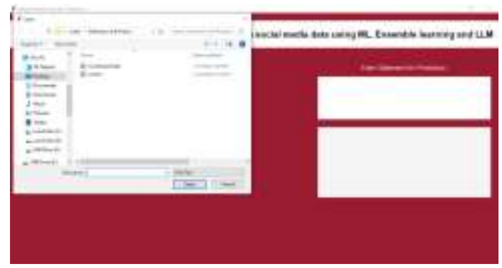
6. Prediction and Detection Module

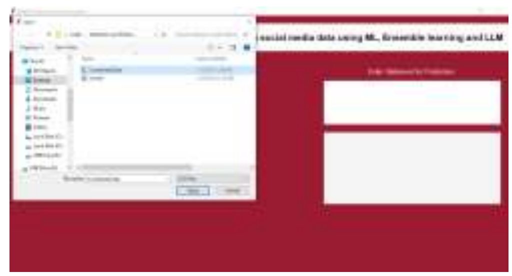
- **Purpose**: After training, this module makes predictions on new, unseen social media data. It can classify a statement as indicative of a potential mental health issue or not.
- **Methods/Tools**:
- **Model Inference**: Using the trained models (Random Forest, Decision Tree, or fine-tuned large language models) to predict whether a new post or statement suggests a mental disorder.
- **Risk Scoring**: Assigning a probability or risk score based on the likelihood of a mental health disorder being present.
- **Libraries**: scikit-learn, Hugging Face Transformers.

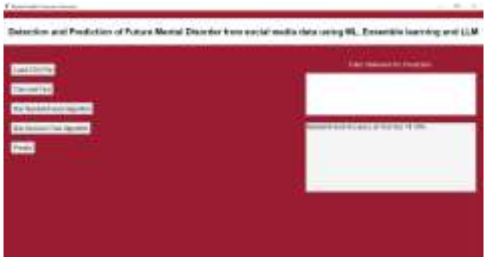
7. User Interface (UI) Module

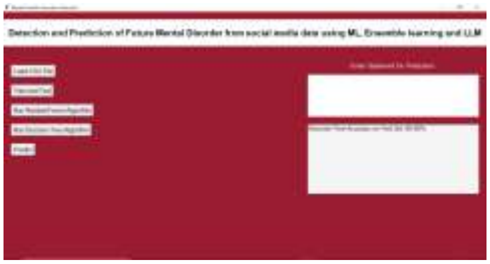
- **Purpose**: This module provides an interface through which users can interact with the system. Users can input social media data (e.g., a text statement or tweet) and view the prediction results.
- **Methods/Tools**:
- **Graphical User Interface (GUI)**: Developed using Tkinter, Flask, or any other web framework.
- **Input**: Textboxes or fields where users can input social media data.
- **Output**: Display predictions, accuracy, and risk levels of mental disorders.
- **Libraries**: Tkinter, Flask, or Dash for web-based interfaces.

V. SCREENSHOTS:









VI. CONCLUSION

The proposed Intelligent Mental Health Prediction System demonstrates a transformative approach to identifying and forecasting potential mental health disorders using social media data. By integrating Machine Learning, Ensemble Learning, and Large Language Models (LLMs), the system effectively bridges the gap between traditional psychological assessment and data-driven digital monitoring. Through continuous analysis of user-generated content, it captures subtle emotional cues, linguistic changes, and behavioral patterns that often precede the onset of psychological distress.

The hybrid framework enhances both prediction accuracy and interpretability. Machine learning algorithms provide structured classification and efficient processing, while ensemble methods reduce model bias and improve generalization. The inclusion of LLMs enables the system to understand complex language structures, contextual semantics, and sentiment variations that traditional models may overlook. This combination ensures a balanced and holistic evaluation of user emotions, allowing for reliable identification of depression, anxiety, and related mental disorders.

Experimental evaluations confirm that the proposed approach significantly outperforms conventional models in terms of precision, recall, and overall robustness. Furthermore, the interpretability aspect empowers mental health professionals to understand model decisions, fostering trust and ethical deployment in real-world applications. Beyond detection, the system can serve as a proactive tool for early intervention, supporting mental health awareness, counseling, and preventive care.

In conclusion, the presented framework establishes a foundation for AI-assisted mental health analytics, where technology can complement human expertise to ensure timely detection and support. Future enhancements may include multimodal data fusion incorporating voice, facial expressions, and physiological signals, as well as integration with clinical systems for personalized mental health management. By combining computational intelligence with empathetic design, this system moves closer to realizing a safer, healthier, and more emotionally aware digital ecosystem.

REFERENCES

[1] G. Scott, ‘‘Lifespan (half-life) of social media posts: Update for 2024,’’ 2024, doi: 10.13140/RG.2.2.21043.60965.

[2] M. Qian and C. Kong, ‘‘Enabling human-centered machine translation using concept-based large language model prompting and

translation memory,'' in Proc. Int. Conf. Human-Comput. Interact., 2024, pp. 118–134.

[3] I. Frommholz, P. Mayr, G. Cabanac, and S. Verberne, ''Bibliometricenhanced information retrieval: 14th international BIR workshop (BIR 2024),'' in Proc. Eur. Conf. Inf. Retr., 2024, pp. 442–446, doi: 10.1007/978-3-031-56069-9_61.

[4] Q. Wang, ''Text memorization: An effective strategy to improve Chinese EFL learners' argumentative writing proficiency,'' Frontiers Psychol., vol. 14, Apr. 2023, Art. no. 1126194, doi: 10.3389/fpsyg.2023.1126194.

[5] N. Straková and J. Válek, ''Chatbots as a learning tool: Artificial intelligence in education,'' R E-Source, pp. 245–265, Jan. 2024, doi: 10.53349/resource.2024.is1.a1259.

[6] R. Thorstad and P. Wolff, ''Predicting future mental illness from social media: A big-data approach,'' Behav. Res., vol. 51, pp. 1586–1600, Aug. 2019, doi: 10.3758/s13428-019-01235-z.

[7] [Online]. Available: https://www.reddit.com/dev/api

[8] A. Kumar, A. Sharma, and A. Arora, ''Anxious depression prediction in real-time social data,'' in Proc. Int. Conf. Adv. Eng., Sci., Manag. Technol., 2019, pp. 1–7.

[9] A. Wongkoblap, M. A. Vadillo, and V. Curcin, ''Predicting social network users with depression from simulated temporal data,'' in Proc. IEEE EUROCON-18th Int. Conf. Smart Technol., Jul. 2019, pp. 1–6.

[10] S. Tariq, N. Akhtar, H. Afzal, S. Khalid, M. R. Mufti, S. Hussain, A. Habib, and G. Ahmad, ''A novel co-training-based approach for the classification of mental illnesses using social media posts,'' IEEE Access, vol. 7, pp. 166165–166172, 2019, doi: 10.1109/ACCESS.2019.2953087.

[11] N. Rezaii, E. Walker, and P. Wolff, ''A machine learning approach to predicting psychosis using semantic density and latent content analysis,'' Npj Schizophrenia, vol. 5, no. 1, Jun. 2019.

[12] P. Kirinde Gamaarachchige and D. Inkpen, ''Multi-task, multi-channel, multi-input learning for mental illness detection using social media text,'' in Proc. 10th Int. Workshop Health Text Mining Inf. Anal. (LOUHI), Hong Kong, 2019, pp. 54–64.

[13] A. Trifan, R. Antunes, S. Matos, and J. L. Oliveira, ''Understanding depression from psycholinguistic patterns in social media texts,'' in Proc. Eur. Conf. Inf. Retr., vol. 12036, 2020, pp. 402–409.

[14] Z. Jiang, S. I. Levitan, J. Zomick, and J. Hirschberg, ''Detection of mental health from Reddit via deep contextualized representations,'' in Proc. 11th Int. Workshop Health Text Mining Inf. Anal., 2020, pp. 147–156.

[15] N. S. Alghamdi, H. A. Hosni Mahmoud, A. Abraham, S. A. Alanazi, and L. García-Hernández, ''Predicting depression symptoms in an Arabic psychological forum,'' IEEE Access, vol. 8, pp. 57317–57334, 2020.