SRIDEVI WOMEN'S ENGINEERING COLLEGE

Department of Computer Science and Engineering(AIML)

(Approved by AICTE and affiliated to JNTUH, An UGC Autonomous Institution, Accredited by NBA &NAAC with A++ Grade

An ISO 9001:2015 certified Institution)

V. N. PALLY(Y), Gandipet(M), Hyderabad-75

2024-2025

DETECTION AND PREDICTION OF FUTURE MENTAL DISORDER FROM SOCIAL MEDIA DATA

MINI PROJECT REVIEW-1

Under the guidance

Dr. Lipika Goel

HOD & Professor

By

G.Ashwitha(22D21A6612)

N.Sharvani(22D21A6624)

P.Preethi(22D21A6627)

ABSTRACT

Detection and Prediction of Future Mental Disorders from Social Media Data. Social media platforms are widely used by individuals to express their emotions, opinions, and mental states. Billions of people around the world use social media daily. Social media platforms, combined with NLP and AI, have greatly simplified many aspects of life. They enable fast communication across countries, allowing people to stay updated with news from around the world in real-time. Additionally, these platforms provide individuals with a space to express their thoughts and emotions through posts and comments. This study utilizes six distinct machine learning models to evaluate their effectiveness in detecting and predicting mental disorders from social media posts. The models employed include Logistic Regression, Decision Tree, Support Vector Machine (SVM), Naïve Bayes, k-Nearest Neighbors (k-NN), and Random Forest. This study utilizes a dataset to automatically detect mental health disorders from social media posts. All of these models were developed, trained, tested, and compared with existing datasets to achieve the most accurate results. In this work the proposed approaches are deployed, compared, and then discussed comprehensively to get the best possible insights and conclusions about the individual's mental health disorders. The approach presented by this models successfully outperforms the number of mental disorders, the variety of models employed and tested, and dataset size used to validate results. The focus is not only on the early detection of existing conditions like depression, ADHD, anxiety, bipolar but also the study is extended to predict successfully potential mental illness that would happen in future.

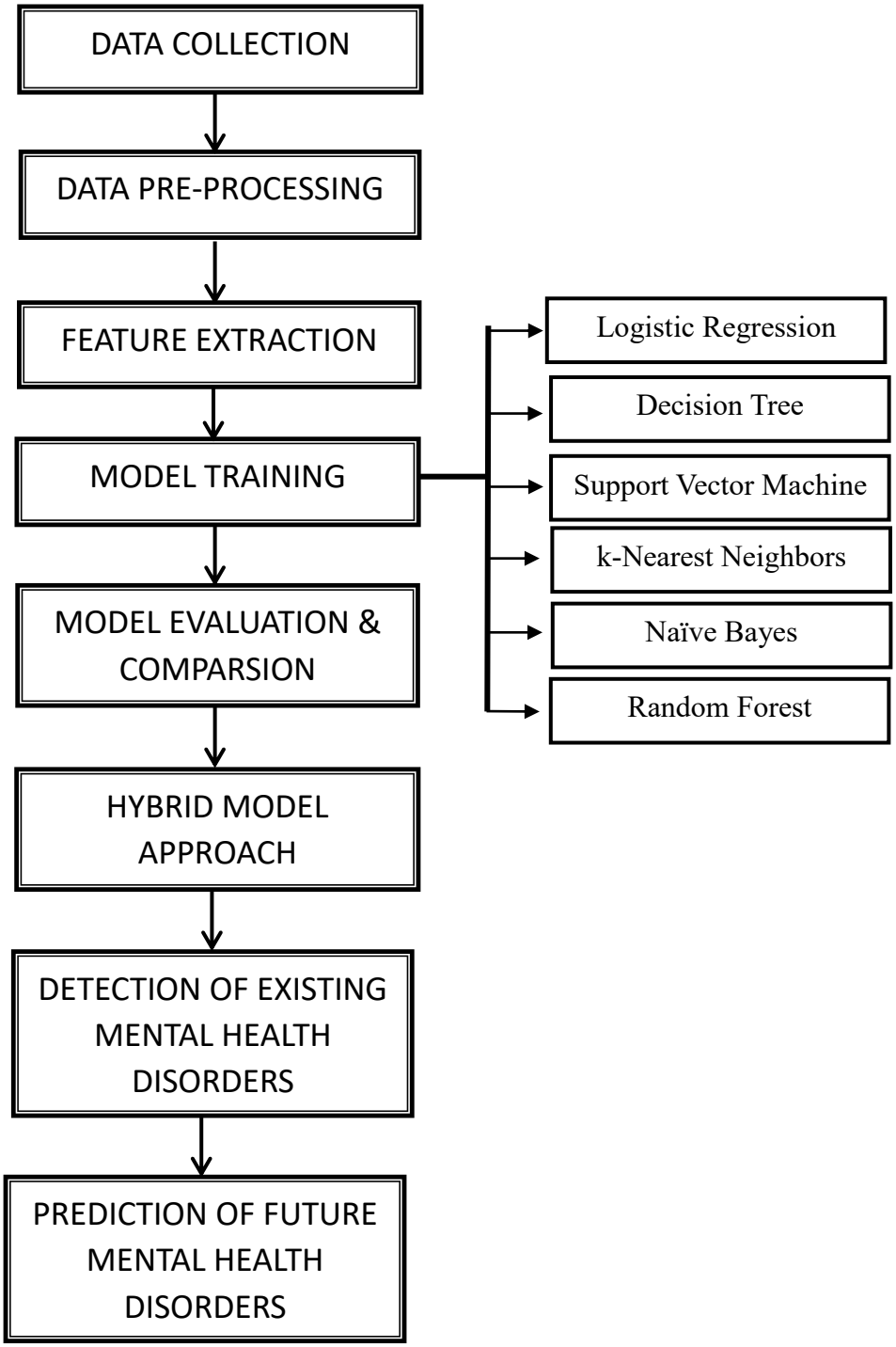
1.PURPOSE

This project aims to develop machine learning models to detect and predict existing mental health disorders, such as depression, anxiety, ADHD, and bipolar disorder, as well as identify new potential mental health conditions from social media posts. Using six models—Logistic Regression, Decision Tree, SVM, Naïve Bayes, k-NN, and Random Forest—the goal is to evaluate their effectiveness in accurate prediction and detection. A hybrid multimodal approach will be created by combining these models to enhance performance. The project seeks to improve early detection methods for both existing and potential mental health disorders, enabling timely support and improved mental health outcomes.

2.SCOPE

The scope of this project focuses on developing machine learning models to detect and predict mental health disorders from social media data. The project begins with collecting and analyzing social media posts, such as those from Twitter, Facebook, and Instagram, to identify emotional expressions and behaviors indicative of mental health conditions. It involves preprocessing the data, including sentiment analysis and feature extraction, to build a dataset for training models. Six machine learning models—Logistic Regression, Decision Tree, SVM, Naïve Bayes, k-NN, and Random Forest—will be implemented to classify and detect disorders such as depression, anxiety, ADHD, and bipolar disorder. The project also explores a hybrid approach, combining the strengths of multiple models to improve prediction accuracy. Additionally, it aims to predict future mental health issues based on current online behavior. Ethical and privacy considerations will be prioritized throughout the process, ensuring user data is respected. Ultimately, the project seeks to enable early detection and intervention in mental health care.

3.MODEL DIAGRAM



4.SYSTEM ANALYSIS

4.1 EXISTING SYSTEM AND DISADVANTAGES

The existing systems for detecting and predicting mental health disorders primarily rely on traditional methods such as manual surveys, face-to-face consultations, and self-reporting questionnaires. These methods are commonly used by mental health professionals to diagnose conditions such as depression, anxiety, and other disorders. In these systems, individuals typically provide information about their mental state through structured interviews or questionnaires, and mental health experts analyze this data to assess the presence of mental health issues. The diagnostic process often involves personal interactions between the patient and the clinician, allowing for an in-depth evaluation of symptoms and behavior.

Disadvantages of the Existing System:

Inaccuracy: Diagnostic methods in traditional systems may not fully capture the emotional fluctuations that occur in daily life, leading to missed or delayed diagnoses.

Time-Consuming: Manual surveys and consultations take time.

Geographical Barriers: People in remote areas might not have access to mental health professionals, limiting timely care.

Social Stigma: The Societal Judgments associated with mental health challenges discourages people from seeking help, impacting both diagnosis and treatment.

Static Data: Existing systems rely on periodic check-ups or self-reports, leading to a lack of real-time, continuous monitoring.

4.2 PROBLEM STATEMENT

Given the limitations of existing systems in the early detection and prediction of mental health disorders, there is a clear need for an advanced, automated system capable of analyzing social media data in real-time. Traditional methods are often slow, have limited reach, and rely on static data, whereas users' online behaviors offer a continuous and rich source of data that can be leveraged for more effective mental health monitoring. The problem lies in detecting mental health disorders promptly and efficiently from social media content, as well as predicting potential future mental health issues based on an individual's behavioral patterns and emotional tone. This system would also address challenges such as stigma, geographic limitations, and inaccurate self-reporting, which all make current diagnostic methods less effective and harder to use widely.

4.3 PROPOSED SYSTEM AND ADVANTAGES

The proposed system aims to use machine learning algorithms to automatically analyze social media posts and predict mental health disorders like depression, anxiety, ADHD, and bipolar disorder etc. By processing data from platforms such as Twitter, Facebook, and Instagram, the system applies Natural Language Processing (NLP) techniques to understand the emotional content of posts. Key components include data collection from public posts, preprocessing for tokenization and sentiment analysis, and feature extraction to identify emotional tone and behavioral patterns. Six machine learning models—Logistic Regression, Decision Tree, SVM, Naïve Bayes, k-NN, and Random Forest—are used for prediction, and their performance is evaluated based on accuracy, precision, recall, and F1-score. A hybrid model will combine the strengths of individual models for better results. The system offers real-time monitoring, early detection of mental health issues, and can reach users globally.

Advantages of the Proposed System:

Early Detection of Mental Disorders: By analyzing social media behavior, the system can identify early signs of conditions like depression, anxiety, and bipolar disorder, enabling early intervention.

Predictive Analytics: The system can predict the likelihood of future mental health issues based on ongoing online behavior, allowing for proactive care.

Wider Accessibility: Social media data is accessible to anyone with an internet connection, making mental health detection possible for underserved populations without in-person consultations.

Real-Time Data Processing: The system continuously analyzes social media posts and comments, providing timely insights and alerts to detect potential mental health concerns quickly.

Confidentiality and Privacy: Users can express their emotions and mental states without revealing their identity, ensuring privacy while allowing the system to detect mental health issues.

Scalable and Efficient: The system can handle large volumes of data from multiple social media platforms, providing insights on a global scale for diverse populations.

Cost-Effective: The system reduces the need for expensive, in-person consultations by utilizing readily available social media data, making mental health monitoring more affordable.

5.FUNCTIONAL & NON-FUNCTIONAL REQUIREMENTS

FUNCTIONAL REQUIREMENTS:

Data Collection: The system must be able to collect data from public social media platforms such as Twitter, Facebook, Instagram, etc., including user posts, comments, and behavioral patterns.

Data Preprocessing and Feature Extraction : The system must preprocess collected data to remove noise and irrelevant information and extract features like sentiment, emotional tone, keywords, and user patterns.

Model Development and Training : The system must implement machine learning models (Logistic Regression, Decision Tree, SVM, Naïve Bayes, k-NN, Random Forest) to analyze and predict mental health disorders based on social media data.

Model Evaluation: The system must evaluate the performance of the machine learning models using metrics like accuracy, precision, recall, F1-score, etc., to compare and identify the most effective model.

Real-Time Monitoring: The system must support real-time monitoring and analysis of social media posts to provide up-to-date insights into a user's mental health status.

Hybrid Approach: The system should combine the outputs of different models into a hybrid approach to increase the accuracy of predictions and make the system more robust.

Data Privacy: The system must ensure user privacy and confidentiality, ensuring that users' identities remain anonymous while analyzing their posts.

Non-Functional Requirements:

Performance: The system should provide real-time predictions and handle large volumes of data with minimal delay, ensuring quick analysis and response.

Scalability: The system should be scalable to handle increasing amounts of social media data and a growing number of users across different platforms.

Reliability: The system should be reliable, with minimal downtime, to ensure continuous monitoring and accurate predictions.

Security: The system must implement robust security measures to prevent unauthorized access to sensitive data and ensure data integrity.

Usability: The system should be easy to use for both developers and mental health professionals, with user-friendly interfaces and clear instructions.

Maintainability: The system should be easy to maintain and update as new machine learning models or social media platforms become relevant.

Compatibility: The system must be compatible with various devices and operating systems to reach a wide audience.

6.HARDWARE & SOFTWARE REQUIREMENTS

Hardware Requirements:

Processor: A multi-core processor (i7 or higher) for efficient data processing and machine learning model training**.**

RAM: At least 16 GB of RAM to handle the large datasets used for training machine learning models.

Storage: A minimum of 500 GB of storage for storing the data, models, and results. SSD storage is recommended for faster read/write operations.

Networking: High-speed internet connectivity to collect real-time social media data from various platforms and transmit data efficiently.

Software Requirements:

Operating System: Windows 10/11 is used for system development and deployment due to its compatibility with most software tools and machine learning libraries.

Programming Languages: Python is the main programming language for machine learning, data analysis, and NLP due to its rich ecosystem and ease of use.

Machine Learning Libraries:

scikit-learn provides implementations of various traditional machine learning models like Logistic Regression, Decision Trees, and Random Forest for predicting mental health disorders.

NLTK (Natural Language Toolkit) is used for processing and analyzing textual data, such as sentiment analysis and text classification.

Pandas and **NumPy** are used for efficient data manipulation, cleaning, and numerical operations on datasets.

Data Collection Tools:

Kaggle and **GitHub** are platforms for acquiring datasets related to social media, mental health, and other relevant domains for training and testing.

Visualization Tools:

Matplotlib and **Seaborn** are used for creating graphical representations of data, such as plots, charts, and heatmaps, to visualize results and trends.

IDE(Integrated Development Environments):

Spyder and **Jupyter Notebook** are integrated development environments that allow writing, testing, and debugging code, with Jupyter supporting interactive data science workflows.