



MENTAL HEALTH PREDICTION SYSTEM USING MACHINE LEARNING TECHNIQUES

¹Nikita Patil, ²Dr.S.K.Wagh

¹Student, ²Professor

¹ Department of Computer Engineering,

¹MES Wadia College of Engineering, Pune,India

Abstract: Mental health disorders such as depression, anxiety, and stress are increasing rapidly due to modern lifestyle changes, academic pressure, and workplace stress. Early identification of mental health conditions is essential for proper treatment and support. This paper presents a Mental Health Prediction System using Machine Learning techniques to predict mental health conditions effectively. The proposed system uses preprocessing techniques, feature extraction, and classification algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine for prediction. The experimental analysis shows that Random Forest achieved the highest prediction accuracy. The developed system can help healthcare professionals and institutions in early mental health assessment.

Index Terms - Mental Health, Machine Learning, Depression Prediction, Anxiety Detection, Artificial Intelligence.

I. INTRODUCTION

Mental health is one of the most important aspects of human well-being and directly affects individual behavior, productivity, and social interaction. In recent years, mental health issues such as depression, anxiety, and stress have increased significantly among students and working professionals. Traditional diagnosis methods are time-consuming and often require specialized healthcare professionals.

Machine learning techniques provide efficient solutions for early prediction and analysis of mental health disorders by identifying patterns in user data. The proposed Mental Health Prediction System aims to predict mental health conditions accurately using machine learning algorithms and healthcare analytics.

II. Literature Review

In recent years, mental health prediction has gained significant attention in the field of machine learning. Many researchers have developed models to detect conditions such as depression, anxiety, and stress using user data. Various studies have applied algorithms like Logistic Regression, Decision Tree, Support Vector Machine (SVM), and Random Forest for prediction. Some systems use survey-based datasets, while others rely on social media or behavioral data.

However, many existing systems face limitations such as lower accuracy, lack of proper data preprocessing, and difficulty in real-world application. Some models also fail to provide early detection effectively.

The proposed system improves upon these limitations by using well-processed data and comparing multiple machine learning algorithms to achieve better prediction accuracy.

III. Problem Statement

Mental health problems are increasing rapidly due to modern lifestyle factors such as academic pressure, work stress, and lack of social interaction. Many individuals do not seek help due to lack of awareness or accessibility to professionals. Traditional diagnosis methods are often slow, expensive, and require trained experts. This creates a gap where early detection becomes difficult.

The main problem addressed in this research is to develop a system that can predict mental health conditions quickly, accurately, and in an accessible way using machine learning techniques.

IV. Objectives

The objectives of this research work are:

- To design and develop a machine learning-based prediction system
- To analyze mental health-related datasets effectively
- To apply and compare multiple machine learning algorithms
- To identify the most accurate prediction model
- To assist in early detection of mental health issues
- To create a system that is easy to use and scalable

V. Methodology

5.1 System Overview

The proposed system is designed to predict mental health conditions using machine learning algorithms. It takes user input data and processes it through trained models to generate predictions.

5.2 Data Collection

The dataset is collected from publicly available sources such as surveys and online datasets. These datasets include various features related to mental health and lifestyle.

5.3 Data Preprocessing

Data preprocessing is an important step to improve model performance.

It includes:

- Handling missing values
- Removing duplicate records
- Encoding categorical variables
- Normalizing numerical data

5.4 Feature Selection

Important features such as age, sleep patterns, stress level, and lifestyle habits are selected to improve prediction accuracy.

5.5 Model Training

Different machine learning algorithms are used:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine

Each model is trained using the dataset and evaluated.

5.6 Prediction Process

After training, the system accepts user input and predicts the mental health condition based on the trained model.

5.7 System Workflow

1. Input Data
2. Preprocessing
3. Feature Selection
4. Model Training
5. Prediction
6. Output Result

VI. Dataset Description

The dataset used in this study consists of various parameters affecting mental health. These include:

- **Demographic Features:** Age, Gender
- **Lifestyle Factors:** Sleep duration, physical activity
- **Work/Academic Factors:** Work pressure, study hours
- **Psychological Factors:** Stress level, mood changes

The dataset is cleaned and structured to ensure better performance of machine learning models.

VII. Implementation / Tools Used

The system is implemented using modern tools and technologies:

- Python: Main programming language
- Pandas: Data manipulation and analysis
- NumPy: Numerical computations
- Scikit-learn: Machine learning algorithms
- Jupyter Notebook: Development environment

The implementation involves loading the dataset, preprocessing data, training models, and evaluating results.

VIII. Results and Discussion

The performance of different models is compared based on accuracy.

Algorithm	Accuracy (%)
Logistic Regression	82%
Decision Tree	85%
SVM	87%
Random Forest	91%

From the results, it is observed that **Random Forest provides the highest accuracy** due to its ensemble learning approach.

The system successfully predicts mental health conditions with good reliability. It also reduces the time required for manual analysis.

IX. Advantages of the System

- Provides early detection of mental health issues
- Reduces dependency on experts for initial screening
- Fast and efficient prediction
- User-friendly system
- Scalable for large datasets

X. Applications

- Hospitals and healthcare centers
- Online mental health platforms
- Corporate wellness programs
- Educational institutions for student support
- Government health initiatives

XI. Future Scope

The system can be further improved in several ways:

- Integration with mobile and web applications
- Use of real-time data for better predictions
- Implementation of deep learning models
- Integration with wearable devices
- Providing personalized mental health suggestions

XII. Conclusion

This research presents a machine learning-based system for predicting mental health conditions. The system uses multiple algorithms to analyze user data and generate accurate predictions. Among all models, Random Forest performed the best in terms of accuracy. The system helps in early detection, which is crucial for proper treatment and mental well-being.

Overall, this research demonstrates that machine learning can play an important role in improving mental health diagnosis and support systems.

References

1. World Health Organization (WHO), “*Mental Health and Well-being*”, Available: <https://www.who.int>
2. K. Shatte, D. Hutchinson, and S. Teague, “*Machine Learning in Mental Health: A Scoping Review of Methods and Applications*”, *Psychological Medicine*, vol. 49, no. 9, pp. 1426–1448, 2019.
3. T. Davenport and R. Kalakota, “*The Potential for Artificial Intelligence in Healthcare*”, *Future Healthcare Journal*, vol. 6, no. 2, pp. 94–98, 2019.
4. S. Bhatia and R. Sood, “*Predicting Depression Using Machine Learning Techniques*”, *International Journal of Engineering Research & Technology (IJERT)*, vol. 7, no. 5, 2018.
5. Scikit-learn Developers, “*Scikit-learn: Machine Learning in Python*”, Available: <https://scikit-learn.org>
6. W. H. Organization, “*Depression and Other Common Mental Disorders: Global Health Estimates*”, WHO Report, 2017.
7. F. Pedregosa et al., “*Scikit-learn: Machine Learning in Python*”, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
8. Kaggle, “*Mental Health Dataset*”, Available: <https://www.kaggle.com>
9. J. Han, M. Kamber, and J. Pei, “*Data Mining: Concepts and Techniques*”, 3rd Edition, Morgan Kaufmann, 2011.
10. I. Goodfellow, Y. Bengio, and A. Courville, “*Deep Learning*”, MIT Press, 2016.