WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Mental Health Detection & Analysis Model Using Machine Learning

Piyush Ahire, Manodip Lohar, Vedant Patil, Ujwal Khairnar, Dr. P.P. Halkarnikar

Department of Computer Engineering, Dr. D. Y. Patil Institute of Engineering Management and Research, Pune, India

A B S T R A C T

In recent study, it has found in the fast-paced modern world, psychological issues like Stress, Anxiety and Depression have been very common among masses. It is crucial to detect mental health condition on a timely basis and cure before it turns into a severe problem. In this paper, we employ Machine learning (ML) algorithms to predict severity levels of depression, anxiety, and stress (DAS) among a diverse group of participants, both employed and unemployed, from various cultural and community backgrounds. Data were gathered using the Depression, Anxiety, and Stress Scale (DASS 42) questionnaire. The psychological states were assessed across five severity levels, utilizing machine learning model known for their high accuracy in predicting mental health issues. During evaluation, the data revealed class imbalances within the confusion matrix, prompting the inclusion of the F1 score metric to more accurately identify the best-performing model, which was the Random Forest classifier. Additionally, the specificity parameter showed the models' heightened sensitivity to detecting negative cases.

Keywords: Mental health detection, Machine learning, Depression prediction, Random Forest, Digital health monitoring

1. Introduction

Mental health issues, especially depression, are skyrocketing in frequency and significantly impacts one's self-care, interpersonal connections, and economic productivity. Depression can vary from slightly troubling psychological symptoms to severe, life-threatening, suicidal tendencies. It is usually caused by pervasive loathing and chronic stress, and certain lifestyle choices that encourage self-neglect. The COVID-19 pandemic has exacerbated these issues, with physical isolation leading to severe social withdrawal, all of which contributed to worsening mental health problems. The World Health Organization (WHO) estimated a 25% increase in the prevalence of depression and anxiety globally owing to the pandemic.

The primary obstacle to addressing mental health concerns is premature recognition, which can result in ineffective solutions. Many people do not understand what their symptoms are showing them, and that adds to their already existing social anxiety. Self-medication or total neglect of mental health issues are as rampant as merited and aggravated over time. Traditional methods of diagnosing clinical interviews and psychological assessments, albeit, useful, tend to be highly subjective, lengthy, and out of reach for a majority of people.

To overcome these problems, our project looks at implementing machine learning, particularly the Random Forest algorithm, for analyzing responses from the DASS-42, which is a 42-item standardized questionnaire that evaluates the severity of depression, anxiety, and stress in an individual. Our objective is to develop a model that predicts depression levels accurately by training it on datasets drawn from the questionnaires' responses.

This methodology has multiple unique benefits: it generates evaluations that are objective and free from biases; it rapidly analyzes large data sets; and it offers real time predictions, which are essential for proactive intervention and fast response actions. Our focus lies in building an inclusive and easy to reach mental health aid system that enables individuals as well as health care professionals to act toward improving mental health proactively. The system can further be developed into an even more advanced intelligent mental health resource with future additions such as Natural Language Processing (NLP).

2. Literature Review

Current mental health detection approaches include self-assessment questionnaires (e.g., PHQ-9), clinical evaluations, and telehealth platforms. However, these methods have limitations, including subjective results, time constraints, and delayed diagnosis. Machine learning has emerged as a promising alternative, with algorithms like Support Vector Machines (SVM), Decision Trees, and Neural Networks used in depression prediction. Among these, the Random Forest algorithm has demonstrated superior performance due to its ensemble learning approach, which enhances accuracy and reliability.

2.1 Predicting Anxiety, Depression, and Stress in Modern Life Using Machine Learning Algorithms

This research focuses on leveraging Machine Learning (ML) algorithms to forecast the presence and severity of mental health conditions such as anxiety, depression, and stress. The study utilizes data obtained through the Depression, Anxiety and Stress Scale (DASS-42) questionnaire, which was

administered to a diverse group of participants including both employed and unemployed individuals from a variety of cultural backgrounds and geographic regions. By incorporating demographic information and self-reported psychological data, the authors were able to train various machine learning models to identify patterns associated with mental distress. The key objective of the paper is to explore the predictive power of these algorithms in identifying early warning signs of mental health issues, thereby enabling more timely and targeted interventions in the broader context of modern lifestyles and occupational challenges.

2.2 Prediction of Depression, Anxiety, and Stress Levels Using DASS-42

In this paper, the authors present a methodology for predicting individual mental health states—specifically depression, anxiety, and stress levels—based on the responses to the DASS-42 questionnaire. The study applies two popular supervised machine learning techniques: Support Vector Machine (SVM) and Logistic Regression (LR). The dataset comprises labeled responses to the DASS-42 items, which are categorized into the three mental health conditions. The models are trained to classify participants into severity categories such as normal, mild, moderate, severe, or extremely severe. The research compares the performance of SVM and LR in terms of accuracy, precision, recall, and F1 score. Results indicate that machine learning offers a viable and automated approach to detecting mental health conditions, which can be especially useful in clinical screening and large-scale mental health assessments.

2.3 Comparative Analysis of Machine Learning Models to Predict Depression, Anxiety, and Stress

This study undertakes a comprehensive evaluation of various machine learning algorithms to determine their effectiveness in predicting mental health conditions, specifically depression, anxiety, and stress (DAS). Using a structured dataset derived from validated psychological assessments such as DASS-42, the research implements and compares the performance of multiple ML models including but not limited to Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Neural Networks. Each model is assessed based on classification metrics like accuracy, precision, recall, and area under the curve (AUC). The paper provides insights into the strengths and limitations of each algorithm when applied to mental health prediction, emphasizing the importance of model selection, data preprocessing, and feature engineering in developing effective psychological assessment tools powered by AI.

3. Methodology

3.1 Data Collection & Preprocessing

The dataset consists of mental health screening responses obtained from standardized tools like HADS and HDRS. Data preprocessing involves handling missing values, normalizing inputs, and selecting key predictive features.

3.2 Machine Learning Model

The model uses the Random Forest algorithm, which employs multiple decision trees to analyze different aspects of participant responses. The final prediction is based on a majority vote among the decision trees.

Key steps in the methodology include:

1) **Feature Selection:** Identifying crucial factors influencing depression risk.
2) **Training and Testing:** Splitting the dataset for model development and evaluation.
3) **Model Performance:** Assessing accuracy, precision, recall, and F1-score.

3.3 System Architecture

The system integrates with digital health platforms for real-time mental health monitoring.

Its components include:

1) **User Interface:** Web or mobile-based input collection.
2) **Prediction Engine:** Machine learning model for depression detection.
3) **Dashboard:** Data visualization and result interpretation.

3.4 Workflow diagram

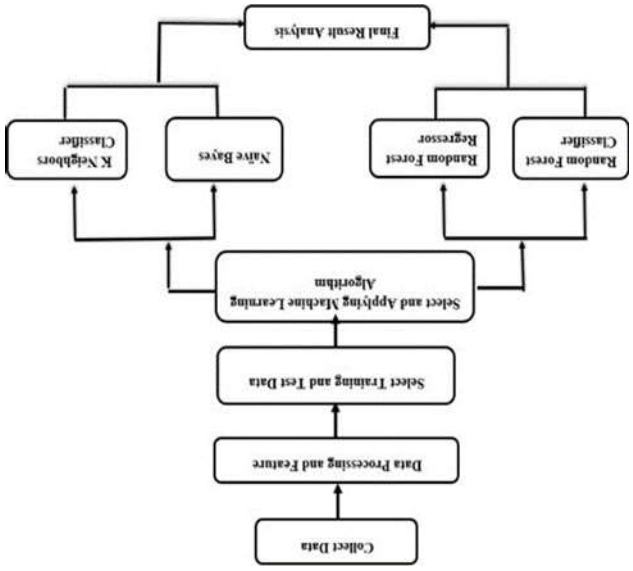

Figure 3.4 Workflow diagram

4. Software Solution And Result

Software Used: NodeJS, ExpressJS, MongoDB, SQL, Python, flask, Pickel

Software Required: VScode Editor, MongoDB compass

Figure 4.1. Homepage

Figure 4.2. Advisory

Figure 4.3. PsyBot

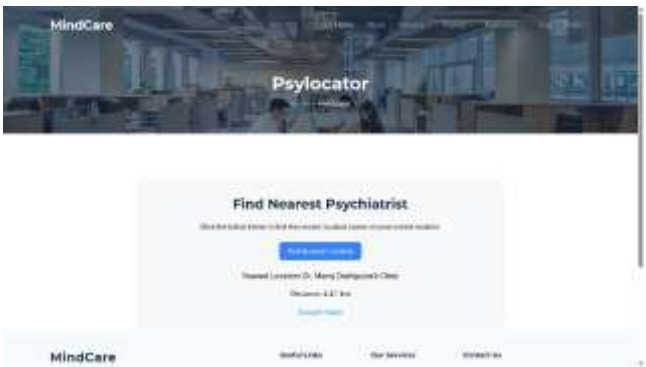

Figure 4.4. PsyLocator

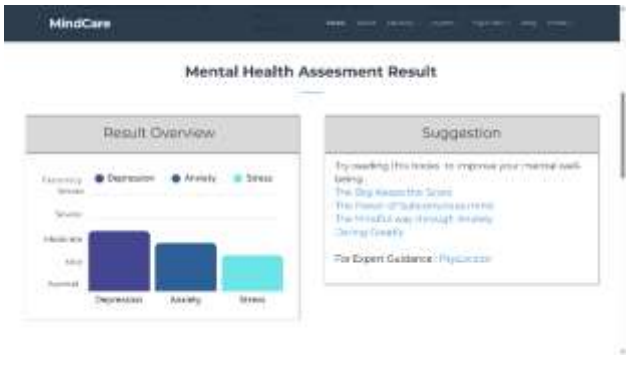

Figure 4.5. TestResult

Figure 4.6. Homepage

5. Results & Discussion

Experiments were conducted on mental health screening datasets, and the Random Forest algorithm demonstrated high accuracy in predicting depression levels.

The findings highlight several key advantages:

1) **Improved Accuracy:** The Random Forest model outperformed traditional diagnostic methods by providing more reliable predictions, thereby enhancing the overall quality of mental health assessments.
2) **Increased Efficiency:** The algorithm was capable of processing large volumes of data at a significantly faster rate, making it suitable for real-time or large-scale mental health screening applications.
3) **Objective Assessment:** By relying on data-driven analysis, the model minimized human biases commonly associated with self-reported questionnaires and subjective evaluations, leading to more consistent and impartial results.
4) **Early Detection:** The predictive capability of the model allowed for the identification of individuals at risk of depression at an earlier stage, enabling timely intervention and potentially reducing the severity of mental health outcomes.

5.1 Result Diagram

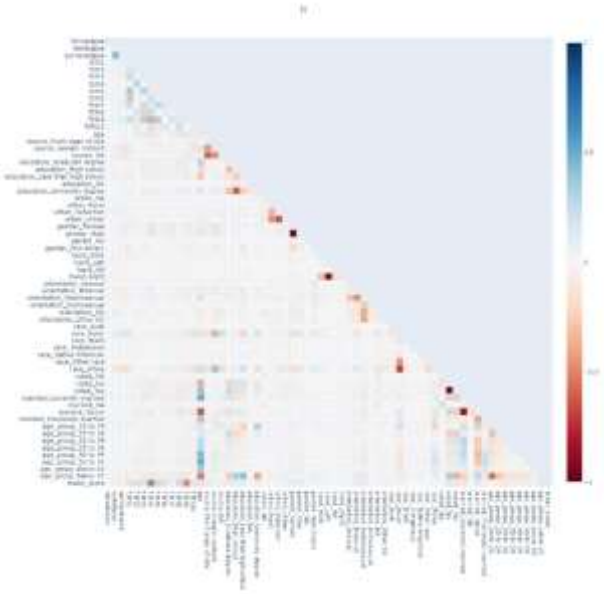

Figure 5.1. Correlation Heatmap

Figure 5.2. Depression, Anxiety and Stress Results from Survey Data

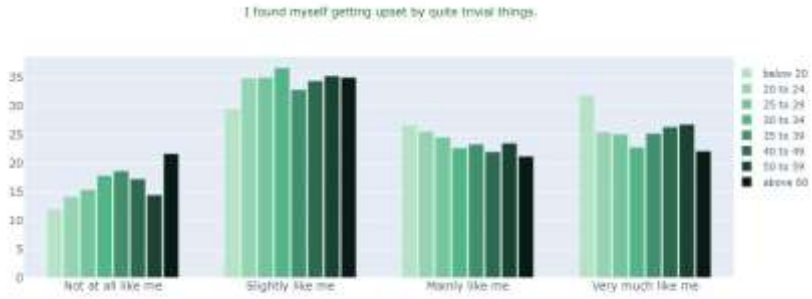

Figure 5.3. Survey Result Difference

6. Applications

The implementation of the Random Forest-based prediction model opens up a wide range of real-world applications that can significantly enhance mental health care and awareness:

1) **Healthcare Support:** The model can assist mental health professionals by offering data-driven insights during diagnosis. By providing an initial analysis of patient responses, it can act as a decision-support tool, helping clinicians identify early signs of depression and reducing diagnostic delays.

2) **Public Health Analysis:** On a broader scale, the model can be used to monitor mental health trends across different regions and demographics. This information can guide policymakers in resource allocation, awareness campaigns, and mental health initiatives at the community level.

3) **Personalized Treatment Plans:** Based on an individual's predicted risk level, the system can help tailor mental health recommendations, such as therapy types, lifestyle changes, or follow-up schedules, ensuring more effective and personalized care.

4) **Telehealth Integration:** The model can be integrated into telemedicine platforms to enhance remote mental health services. It enables real-time assessments and continuous monitoring, making mental health support more accessible, especially in underserved or rural areas.

7. Future Scope

This research lays the foundation for a range of future developments aimed at making mental health prediction models more adaptive, intelligent, and accessible.

Key directions for advancement include:

1) **Personalized Mental Health Models:** Future versions of the system can be designed to adapt predictions based on individual characteristics such as age, gender, medical history, lifestyle, and environmental factors. This personalization would lead to more accurate and meaningful mental health assessments tailored to each user's unique profile.

2) **Real-Time Monitoring:** By integrating the model with wearable health devices—such as smartwatches and fitness trackers—it would be possible to continuously monitor physiological signals (e.g., heart rate, sleep patterns, and activity levels). Real-time data could enhance the system's ability to detect early warning signs of mental distress and trigger timely alerts or recommendations.

3) **Telehealth Expansion:** incorporating deep learning techniques such as neural networks could significantly boost the model's accuracy and capability. These advanced algorithms can uncover complex, non-linear relationships in data, enabling more nuanced understanding and prediction of mental health conditions.

8. Conclusion

This research introduces a machine learning-based approach for predicting depression levels using data from the DASS-42 questionnaire, with the Random Forest algorithm serving as the core predictive model. The algorithm demonstrated high accuracy, scalability, and resilience to overfitting, making it a suitable choice for analyzing complex mental health data. Its ability to identify the most relevant features contributing to depressive symptoms allows for more precise and personalized evaluations. The findings suggest that machine learning can be a valuable tool in augmenting traditional psychological assessments, offering objective, data-driven insights into mental health conditions.

In real-life applications, this system has the potential to revolutionize early mental health diagnostics by making screening more accessible and proactive. For instance, it can be integrated into mobile health apps, enabling individuals to monitor their emotional well-being privately and regularly, without needing immediate clinical access. This is particularly impactful in regions where mental health resources are limited or stigmatized. Additionally, educational institutions, workplaces, and healthcare providers can use such models for early identification of individuals at risk, allowing for timely support and intervention. By bridging the gap between technology and mental healthcare, the proposed approach contributes to a more preventive, inclusive, and responsive mental health ecosystem.

References

1. S. Singh, H. Gupta, P. Singh and A. P. Agrawal, "Comparative Analysis of Machine Learning Models to Predict Depression, Anxiety and Stress," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 1199-1203, doi: 10.1109/SMART55829.2022.10047752. keywords: {Support vector machines;Pandemics;Anxiety disorders;Mental health;Predictive models;Depression;Market research;Depression;Anxiety;Stress;Loneliness;Health;Machine Learning;Esemble Learning},

2. 2. K. S. Srinath, K. Kiran, S. Pranavi, M. Amrutha, P. D. Shenoy and K. R. Venugopal, "Prediction of Depression, Anxiety and Stress Levels Using Dass-42," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-6, doi: 10.1109/I2CT54291.2022.9824087. keywords: {Support vector machines;Anxiety disorders;Human factors;Mental health;Depression;Prediction algorithms;Classification algorithms;ML;DAS;DASS;DASS42;SVM;LR},

3. Anu Priya, Shruti Garg, Neha Prerna Tigga, Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms, Procedia Computer Science, Volume 167, 2020, Pages 1258-1267, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.03.442. (https://www.sciencedirect.com/science/article/pii/S1877050920309091)

   Abstract: In the fast-paced modern world, psychological health issues like anxiety, depression and stress have become very common among the masses. In this paper, predictions of anxiety, depression and stress were made using machine learning algorithms. In order to apply these algorithms, data were collected from employed and unemployed individuals across different cultures and communities through the Depression, Anxiety and Stress Scale questionnaire (DASS 21). Anxiety, depression and stress were predicted as occurring on five levels of severity by five different machine learning algorithms – because these are highly accurate, they are particularly suited to predicting psychological problems. After applying the different methods, it was found that classes were imbalanced in the confusion matrix. Thus, the f1 score measure was added, which helped identify the best accuracy model among the five applied algorithms as the Random Forest classifier. Furthermore, the specificity parameter revealed that the algorithms were also especially sensitive to negative results.

   Keywords: Decision Tree(DT); K-NN; Naïve Bayes(NB); Random Forest Tree(RFT); Support Vector Machine(SVM)

4. IAES International Journal of Artificial Intelligence (IJ-AI)s