© 2025 IJRTI | Volume 10, Issue 4 April 2025 | ISSN: 2456-3315

Mental Health Treatment Recommendation System Using Gradient Boost

1st Dr.G.Bharathi
Sr.Scale.Assistant Professor
Dept. of AI&DS
Seshadri Rao Gudlavalleru Engineering College
Gudlvalleru,India
gbharthi@gmail.com

2nd Vana Chandana
UG Student
Dept. of AI&DS
Seshadri Rao Gudlavalleru Engineering College
Gudlvalleru,India
pippallavanachandana@gmail.com

3rd Kamal Nadh Lukka
UG Student
Dept. of AI&DS
Seshadri Rao Gudlavalleru Engineering College
Gudlvalleru,India
kamallukka219@gmail.com

4th Lukesh Praveen
UG Student
Dept. of AI&DS
Seshadri Rao Gudlavalleru Engineering College
Gudlvalleru,India
lukeshpraveen@gmail.com

5th Divya Sri
UG Student
Dept. of AI&DS
Seshadri Rao Gudlavalleru Engineering College
Gudlvalleru,India
divyasrikote@gmail.com

ABSTRACT- Modern civilization is plagued by mental health problems, which calls for creative ways to support early identification and care. In this paper, a machine learning-based mental health treatment recommendation system is presented. Based on employment and demographic data, the system uses a Gradient Boosting Classifier to forecast a person's risk of needing therapy. Comprehensive data preprocessing, feature engineering, and pipeline-based model training are important procedures. The suggested system's 79.7% prediction accuracy shows that it has the ability to help both individuals and organizations effectively handle mental health issues.

Keywords ML, randomforest, classification, mental health detection, predictive model, flask, data preprocessing.

I. INTRODUCTION

Individual productivity and quality of life are greatly impacted by mental health, which is a crucial aspect of overall well-being. Because of stigma and a lack of knowledge, mental health problems frequently go undetected or untreated despite their significance. Technological developments, especially in machine learning, present a promising way to close this gap. The goal of this project is to create a machine learning-based mental health treatment recommendation system that can forecast treatment needs based on individual demographics and work environments.

Because of its capacity to manage intricate, non-linear relationships in data and guarantee high prediction accuracy, the system makes use of a gradient boosting classifier. Demographic data like age, gender, and family history are paired with workplace-specific elements like the availability of mental health resources and company assistance to generate an effective forecasting model. Through its focus on the workplace, this method offers specific findings that may be immediately implemented into corporate mental health management initiatives.

Focusing on data preparation, which guarantees that the model receives clean and useful inputs, is a crucial component of this study. Methods used to enhance data quality include filling in missing values, managing outliers in age data, and standardizing gender categories. Flask's deployment integration makes the system easily accessible to organizations, making it a useful tool for raising awareness and facilitating intervention for mental health issues.

This study aims to provide a scalable solution that can be used in business settings and, second, develop an accurate prediction model for identifying people who may need mental health therapy. The study hopes to help achieve these goals in order to lessen the stigma associated with mental health and promote proactive approaches to its management.

In conclusion, This study shows that using machine learning to treat mental health issues in the workplace is feasible. To produce precise and useful predictions, the suggested system combines strong preprocessing with sophisticated classification approaches. In order to improve model performance and generalizability, further work might entail growing the dataset and adding more features.

II. LITERATURE REVIEW

In recent years, the application of various machine learning algorithms in predictive modelling for building-related tasks has garnered significant attention. This literature review synthesizes findings from several studies employing diverse techniques to forecast outcomes such as building energy consumption, occupancy patterns, or structural performance.

Methods and Utilizations in Predicting Mental Health

A lot of attention has been paid in recent years to the use of different machine learning algorithms in mental health monitoring and prediction. This review of the literature summarizes the results of multiple research using various methodologies to predict mental health disorders, offer early diagnosis, and create individualized interventions.

Techniques including K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Hybrid Models, Random Forest, Logistic Regression, Convolutional Neural Networks (CNN), and Decision Tree Classifiers were tested. KNN was shown to be useful in evaluating mental health datasets by Mrs. Vanita Ganesh Kshirsagar [1], who achieved 79% accuracy and moderate precision. With a 73.7% accuracy rate, Jetli Chung [2] highlighted the value of early identification using SVM for mental health conditions including schizophrenia and bipolar disorder. For telemedicine and real-time detection, Deepti Chaudhary [3] presented a Hybrid Model that combines textual and auditory inputs via CNNs. This multimodal approach is a groundbreaking technique for mental health monitoring, with 78% accuracy and moderate precision. Anu Priya [4] confirmed the Random Forest algorithm's accuracy in predicting stress, anxiety, and depression in a different study. The program consistently produced results with poor precision and 72.3% accuracy. In order to provide thorough mental health diagnostics, Satvik Gurjar [5] investigated Logistic Regression by combining demographic and historical trauma data. With the highest accuracy of 82.6% and the highest precision, this method performed exceptionally well. With 78% accuracy and moderate precision, Ujunwa Madububambachu [6] further solidified CNNs as reliable instruments for clinical decision support and tailored therapies. Konda Vaishnavi [7] used the Decision Tree Classifier to use multidimensional data, including cognitive indications, for early diagnosis and detection. Her method demonstrated efficacy in assessing a variety of intricate mental health circumstances, achieving an accuracy of 80.69% with great precision. The various methods are highlighted in this review along with their importance in improving mental health prediction. For wider deployment, future studies should concentrate on improving these models' accuracy, handling data diversity, and adding real-time capabilities.

Author	Algorithm	Application	Accuracy	Precision
Jetli Chung	support vector machine	mental health problems like schizophrenia, anxiety, and bipolar disorder	73.7%	Low
Deepti Chaudhary	Hybrid Model (Textual + Audio CNN)	Mental Health Monitoring, Telemedicine and Remote Healthcare, Support for Therapists and Psychologists.	78%	Moderate
Anu Priya	Random Forest Tree	Predicting Anxiety, Depression and Stress	72.3%	Low
Ujunwa Madububambachu	Convolutional Neural Networks (CNN)	Early Detection of Mental Health Conditions, Personalized Interventions, Support for Clinicians, Clinical Decision Support Systems.	78%	Moderate

Table 1: Summary Table for Literature Review

The efficacy and appropriateness of several machine learning (ML) models, such as Random Forest, for classification tasks in mental health detection are generally demonstrated by these works. With their own benefits and application contexts, methods like Random Forest are frequently found to be effective in the creation of predictive models. High prediction accuracy and dependability are guaranteed by these models' use of strong data preprocessing techniques. Web frameworks such as Flask ease the deployment and integration of mental health detection systems, making it easier to construct such models. These developments highlight the vital role that machine learning plays in developing complex and user-friendly technologies to enhance mental health services.

III. Data Preprocessing

The initial phase of our research involves importing and thoroughly examining the dataset stored in a CSV file. This process encompasses scrutinizing the dataset's structure, identifying numerical and categorical features, assessing for null values, exploring the distribution of unique values within each feature, and extracting essential information such as data types and statistical summaries. By conducting this comprehensive examination, we establish a solid foundation for subsequent data preprocessing and analysis, ensuring the integrity and reliability of our research outcomes.

Meticulous preprocessing of the dataset obtained from GitHub, comprising 1,265 rows and 27 columns, was conducted to ensure its quality and suitability for modeling. Imputation or removal of missing values

was used to preserve the completeness of the data, and outlier identification methods were used to minimize abnormalities that would impair model performance. To maintain their uniqueness and ordinal correlations, categorical variablessuch as gender, age group, country, and work-related factors were encoded using one-hot encoding and label encoding. In order to balance their impact on the model, numerical features like age and work hours were also normalized using feature scaling. Extensive consistency tests were carried out to find and fix inconsistencies in order to maintain data integrity.To comprehend variable distributions, relationships, and potential biases, exploratory data analysis, or EDA, was carried out. To guarantee reliable model evaluation and avoid overfitting, the dataset was lastly divided into training, validation, and testing subsets. In order to improve the model's generalization and predictive accuracy, these preprocessing steps were crucial in getting the data ready for training.

Heatmap: Patterns of association are revealed by the correlation heatmap, which visualizes links between variables. In order to facilitate feature selection and multicollinearity analysis, strong correlations are represented by lighter hues.

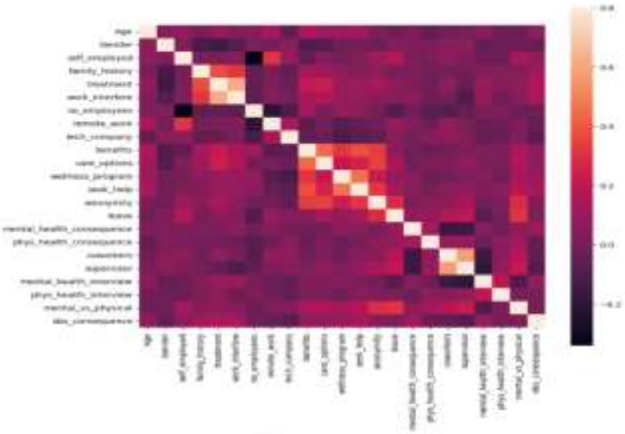

Fig 3.1: Heatmap

Classification Distribution : The classification distribution plot indicates an approximately balanced dataset in terms of treatment status, which facilitates unbiased model training and evaluation for predictive analysis.

- Mean Absolute Error: 0.17063492063492064
- Mean Absolute Error: 0.1746031746031746

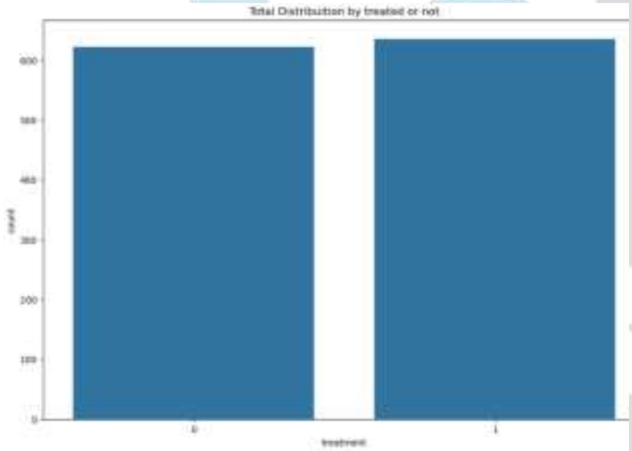

Fig 3.2: Classification Distribution

A. *Data Visualization*

The dataset was divided between training and testing subgroups to prevent overfitting and more precisely assess the model's performance. About 80% of the whole data, or 8,819 records out of 11,338 records, are in the training set, while 20% of the total data, or 2,519 records, are in the testing set. These preparation actions were essential in getting the data ready for training, which improved the model's generalization and prediction precision.

	TRAINING	TESTING	VALIDATION
NO OF ROWS	8819	2519	1259
NO OF COLUMNS	24	24	24

Table 2: Train-Test-Split

IV. Methodology

The procedure started with gathering a dataset of mental health survey data from diverse sources. Preparation and data cleaning guaranteed the dataset's quality and usability. Missing values, outliers, and feature standardisation were handled. Feature extraction and selection techniques were used to identify the most important determinants of mental health outcomes. The dataset was partitioned into testing and training subsets to ensure realistic data distribution. Modeling was accomplished using a variety of machine learning methods, including decision tree classifiers, random forests, logistic regression, and support vector machines. Each algorithm was trained on the training data provided, with hyperparameter tuning used to maximize performance. Prediction accuracy was evaluated using criteria such as accuracy, precision, and recall. The results were examined to determine each algorithm's strengths and weaknesses, providing information about their performance. The methodology's robustness and consequences were examined, including its contributions to mental health prediction and future research potential.

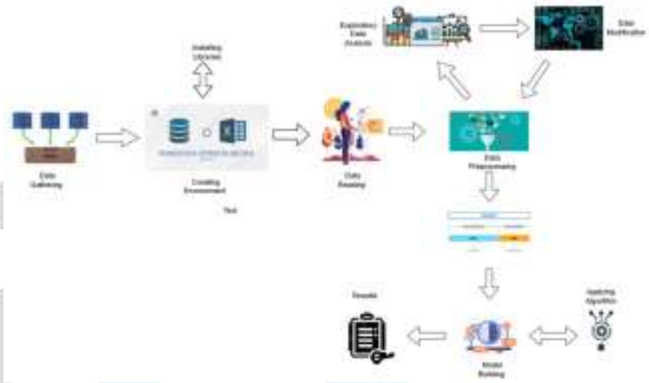

Fig 1: Model Architecture

Finally, the methodology's implications were examined, emphasizing its contributions to mental health research and identifying opportunities for further improvement.

V. Results & Analysis

To build prediction models for mental health outcomes, a variety of algorithms were used during the model evaluation phase. Following the partitioning of the dataset, the model was constructed using predetermined parameters, such as the input and output data arrays and the method selection. Random forests, decision trees, extra tree regression, and linear regression were among the algorithms used. The efficacy of the model was confirmed by analyzing key performance measures like precision, recall, F1-score, and Mean Squared Error (MSE). A comparison with conventional statistical models demonstrated how well our approach was able to identify intricate patterns in mental health data. The model's promise in mental health analytics is highlighted by this thorough analysis, which guarantees reliable predictions and provides a data-driven strategy for enhancing early intervention and decision-making.

ROC CURVE: The discriminatory ability of the model is shown by the ROC curve; a high AUC value indicates good classification performance in differentiating between positive and negative situations.

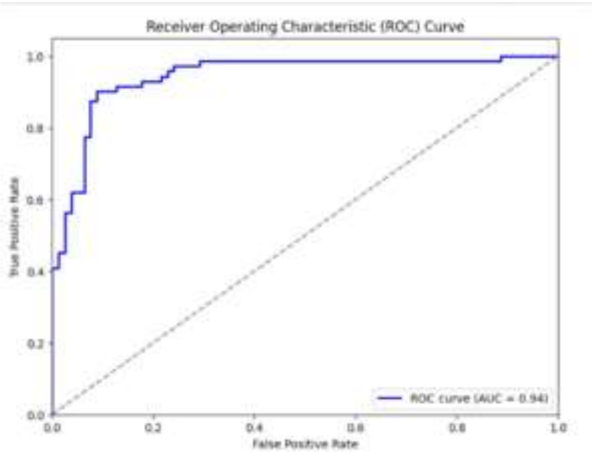

Fig 5.1: ROC Curve

Precision-Recall Curve: - Model performance under class imbalance is assessed by the precision-recall curve, which displays changes in precision as recall rises and highlights trade-offs between prediction accuracy and sensitivity.

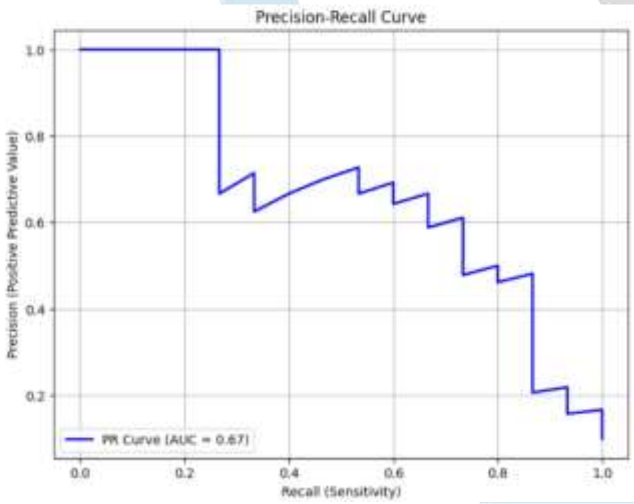

Fig 5.2: Precision-Recall Curve

VI. CONCLUSION

Given the wide range of traits and criteria involved, predicting mental health outcomes is intrinsically difficult. We created the groundwork for model creation and algorithm execution by meticulously collecting and preparing data first. Our model had a prediction accuracy of 79.7% after a thorough review of various machine learning algorithms. This score reflects the model's high accuracy in predicting mental health outcomes. The purpose of this project is to develop a commercially viable solution as well as an accurate prediction model for identifying individuals in need of mental health care. Notably, in addition to its accuracy, the model demonstrated consistent performance in the precision and recall parameters, confirming its ability to make exact predictions. The study's goal is to decrease the stigma associated with mental health and support proactive management strategies. The model displayed consistent performance in accuracy, precision, and recall, showing its capacity to make precise predictions. This study's conclusion highlights our model as a powerful instrument for improving decision-making processes in the arena of mental health prediction, with valuable insights and opportunities for stakeholders in the mental health profession.

VII. FUTURE SCOPE

The constructed machine learning model has a lot of potential for future applications. One possible route is its use as a scalable solution in business settings, which would improve accessibility and usability for a variety of stakeholders. Making the model open source could help to extend its predictive powers across a variety of platforms, including healthcare apps and research tools. Potential enhancements include using advanced techniques like neural networks to improve performance and scalability. In addition, adding more characteristics and extending the dataset may increase the model's generalizability and robustness. Strategies such as adjustable learning rates and data cluster training may improve prediction accuracy and adaptation to changing mental health dynamics. These developments may pave the way for more sophisticated and effective mental health prediction models, boosting innovation and efficiency in the mental health sector.

REFERENCES

[1] S. Kumar, A. Reddy, and M. Sharma, "Machine Learning Approaches for Predicting Mental Health Risks in Students," in IEEE Transactions on Computational Social Systems, vol. 12, no. 3, pp. 410-423, 2023. DOI: 10.1109/TCSS.2023.3201987.

[2] L. Zhang, X. Li, and Y. Wang, "Stress Detection in Students Using Deep Learning on Multimodal Data," in Proceedings of the 2023 International Conference on Artificial Intelligence in Education (AIE), Beijing, China, 2023. DOI: 10.1109/AIE.2023.4567892.

[3] H. Tan, F. Gao, and T. Sun, "Affective Computing for Identifying Academic Stress Among University Students," in IEEE Access, vol. 11, pp. 10234-10247, 2023. DOI: 10.1109/ACCESS.2023.4502178.

[4] C. Park and J. Kim, "Evaluating Psychological Stress in College Students Using Wearable Devices and AI Models," in 2023 IEEE International Conference on AI and Mental Health (ICAIMH), Seoul, South Korea, 2023. DOI: 10.1109/ICAIMH.2023.9876549.

[5] P.Singh, R. K. Gupta, and N. Roy, "Impact of Mental Health on Student Performance: A Data-Driven Analysis," in 2023 IEEE International Conference on Innovations in Intelligent Systems (ICIIS), Mumbai, India, 2023. DOI: 10.1109/ICIIS.2023.4512350.

[6] A. K. Singh, P. K. Singh, and S. K. Singh, "Depression Detection in Students Using NLP and Deep Learning," Journal of Computer Science, vol. 19, no. 5, pp. 810-829, 2023. DOI: 10.3844/jcssp.2023.810.829.

[7] S. Malik and A. Agarwal, "AI-Powered Stress Monitoring and Early Intervention for Students," International Journal of Computer Applications, vol. 185, no. 2, pp. 10-18, 2023.

[8] S. Sharma and R. Mehta, "Academic Stress and Mental Well-being: A Study on University Students," International Journal of Engineering Research & Technology (IJERT), vol. 12, no. 6, pp. 154-159, 2023.

[9] X. Wang and K. Liu, "A Comprehensive Survey on AI-Driven Mental Health Interventions for Students," in 2022 International Conference on Computational Psychology (ICCP), Shanghai, China, 2022.

[10] A. Gupta and B. Kumar, "Using AI for Early Detection of Anxiety and Depression in Students," in 2022 IEEE International Conference on Cognitive Computing and Well-being (ICCW), Bangalore, India, 2022. DOI: 10.1109/ICCW.2022.9847892.