

Depression Prediction System Using Machine Learning for Student Mental Health Analysis

Mr. D. Ashok

Assistant Professor

Department of Computer Science & Artificial intelligence Central University of Andhra Pradesh,
Ananthapuramu, India

Email : ashokdhakuri@cuap.edu.in

O. Nalina

Department of Computer Science & Artificial intelligence Central University of Andhra Pradesh,
Ananthapuramu, India

Email : nalinaobaiah@gmail.com

Abstract

With the rise in the number of issues like unhealthy sleeping routine, stress, and the financial problems faced by the students, student mental health has been evolving as an important concern. It directly affects a student's studies and their ability to concentrate on their work, and not just their mental state. This study presents a ML based Depression Prediction System for efficient recognition of students prone to early depression.

The ML system takes into account the students' features that are relevant and include Gender, Age, study satisfaction, Academic pressure, Duration of sleep, eating habits, Duration of study, financial problems, suicidal thoughts, mental disorder in family history. After processing the attributes such as one-hot encoding and standard scaling, the data is made suitable for algorithms and is ready to increase the system's accuracy.

Numerous ML based classification models like Logistic Regression, SVM, KNN, Random forest, Gradient boosting and Extra trees classifier have been taken into account and tested using Accuracy, F1-score, Recall, Precision, and Confusion Matrix. All models were trained and tested and the model with the best performance has been selected for depression prediction.

A Gradio based UI has been implemented which aids in uploading datasets, training models, visualizing results, and making a prediction of the given data both individually and in batches. At the end, a prediction of Low, Medium or High depression risk will be provided.

The proposed system is expected to be useful in recognizing students with mental health problems and making them aware of their health concerns. However, this system can in no way substitute a proper medical examination.

Keywords : Machine Learning, Depression Prediction, Student Mental Health, Classification, Random Forest, Support Vector Machine, K-Nearest Neighbor, Ensemble Learning, Predictive Analytics, Data Preprocessing, Gradio Interface, Artificial Intelligence, Mental Health Analysis, Student Well-being, Healthcare Analytics.

I. INTRODUCTION

Mental health is of critical importance for student learning as well as for students' overall well-being. One of the most prevalent mental disorders amongst students today is depression. Student depression is frequently linked to excessive academic pressure, finances, insufficient sleep, an unbalanced lifestyle, and emotional stress. The early diagnosis of depression can allow educational institutions to offer their students appropriate counseling and immediate assistance.

Machine learning is efficient at determining underlying behavioral patterns in students and forecasting potential mental health issues. The present research work, 'Depression Prediction System' involves utilizing machine learning algorithms that can determine potential depression in students based on several factors including academic pressure, student satisfaction towards their studies, sleep length, financial concerns, food habits and genetic tendency to mental illness. Different classification algorithms are used and their effectiveness is evaluated so that the best performing model can be selected. A user-friendly and effective Gradio-based interface has been created for easier interaction and real-time predictions.

II. LITERATURE REVIEW

Depression detection is becoming an important research area as students have difficulty concentrating on their studies and social interactions are affected, including their mood and emotional state. Manual methods such as traditional counseling and questionnaires can take up much time and effort for mass student communities. Machine learning techniques solve these problems through analyzing abundant data about students in order to discover the hidden pattern correlated to the depression state [1],[3]. Previous studies confirmed that accuracy of depression prediction can be improved by using academic, behavioral and psychological features.

Several ML algorithms such as Logistic Regression, SVM, KNN, Random Forest, Gradient Boosting, and Extra Tree Classifier are widely adopted for mental health prediction systems. Among them, ensemble learning methods such as Random Forest and Gradient Boosting yield accurate prediction for non-linear data because they can catch both nonlinear relationship in between data. This model is very commonly used in healthcare prediction systems due to the robustness and reliability [2],[6]. It is also observed in studies that Academic pressure, Sleep duration, Study satisfaction, financial difficulties, Eating habits and family history of mental health problem are most sensitive variables. This psychological condition is largely associated with bad sleep, heavy course load, unhealthy life style and emotional stress [3],[5]. Thus, using these features in ML models enhances the model prediction accuracy.

Recent trends on Explainable Artificial Intelligence (XAI) are important in mental health prediction systems due to its benefits of transparency and fairness. Users will have a general idea of the decision making process that influenced the prediction based on different input parameters. This approach would improve the trust, confidence, and useability of the ML systems on healthcare-related applications. Latest developed mental health prediction models are also supported by the visualization tools, and real-time prediction functions as well as human friendly interfaces [6],[7].

Despite these improvements, there are still limitations of the existing system. It relies mostly on small size data, Lack real time implementation/practical applications and manual input system. Limited diversity of feature and data imbalance are also issue. In order to overcome this issue, the system applies variety machine learning models, the model comparison and visualization technique along with a Gradio based interface [2],[6].

II. METHODOLOGY

The process for the proposed system: data collection, data preprocessing, training model, evaluating the model, and making the prediction.

Data Preprocessing

Imputing for missing value.

Encoding categorical value into numerical using One-Hot Encoding.

Scaling numerical value using StandardScaler.

Split the dataset into training and test sets.

Machine Learning Algorithms Used

We use the following models in this project.

Logistic Regression

Support Vector Machine (SVM)

K-Nearest Neighbor (KNN)

Random Forest

Gradient Boosting

Extra Trees Classifier

Evaluation Metrics

We use accuracy, precision, recall, F1-score and Confusion matrix to measure and evaluate our models so that we can choose the best model to predict depression.

IV. RESULTS

This paper presented a Depression Prediction System developed by applying a set of ML techniques for students depression risk prediction. Various techniques like one-hot encoding, standard scaling were utilized in the preprocessing steps for increasing model performance and accuracy. We tested several traditional and ensemble ML models like Logistic Regression, SVM, KNN, Random Forest, Gradient Boosting and Extra Trees Classifier and evaluated their performance using different metrics.

The results of the experiment showed that the ensemble models outperformed than traditional ML classifiers. From all the model used, the Random Forest and Extra Trees Classifier demonstrated excellent accuracy results because the two model is good in dealing with complicated relationships between features and also can effectively handle the non-linear patterns of dataset. It provided a better classification in terms of students depression risk prediction.

The performance evaluation of the models was performed using the parameters of Accuracy, Precision, Recall, F1-Score, Confusion Matrix. From the confusion matrix, it can be seen that the model correctly predicted depressed and non-depressed students without any misclassification and both Precision and Recall demonstrated a well performance of identifying the depressed students.

Feature importance test displayed that Academic pressure, Sleep time, Financial issue, Work and study satisfaction, Suicidal tendency are among the most important factor for depression prediction. High academic pressure, bad sleep habit, poor lifestyle, and low financial capacity tend to result in a higher possibility for students depression risk. These factors prove that the academic, behavior and mental status of students influence the prediction.

A GUI using Gradio was also developed in order to make the system easy to use and friendly interactive. It enabled user to upload the dataset, train ML models and visualizing performance measures, and performing Manual prediction of

depression risk score. The system can accurately classify students depression risk levels into Low, Medium and High with probability in the output.

To conclude, the experimental result shows that machine learning technology is effectively applied in predicting students depression risk at the early stages. The system obtained a reliable prediction, a good usability and acceptable performance, it can be used in universities or for mental awareness campaigns.

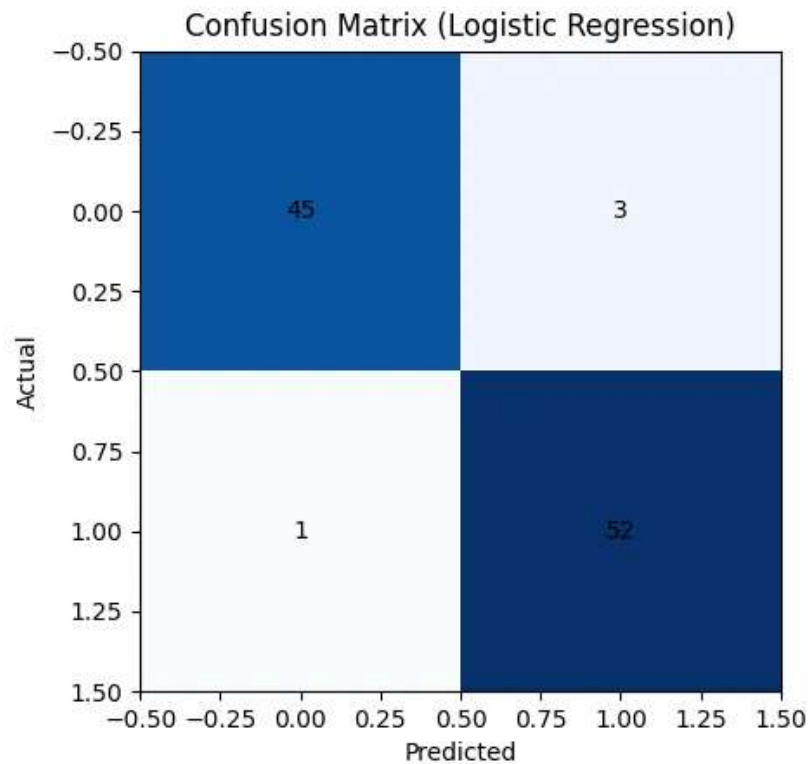


Fig. No.1: Confusion Matrix

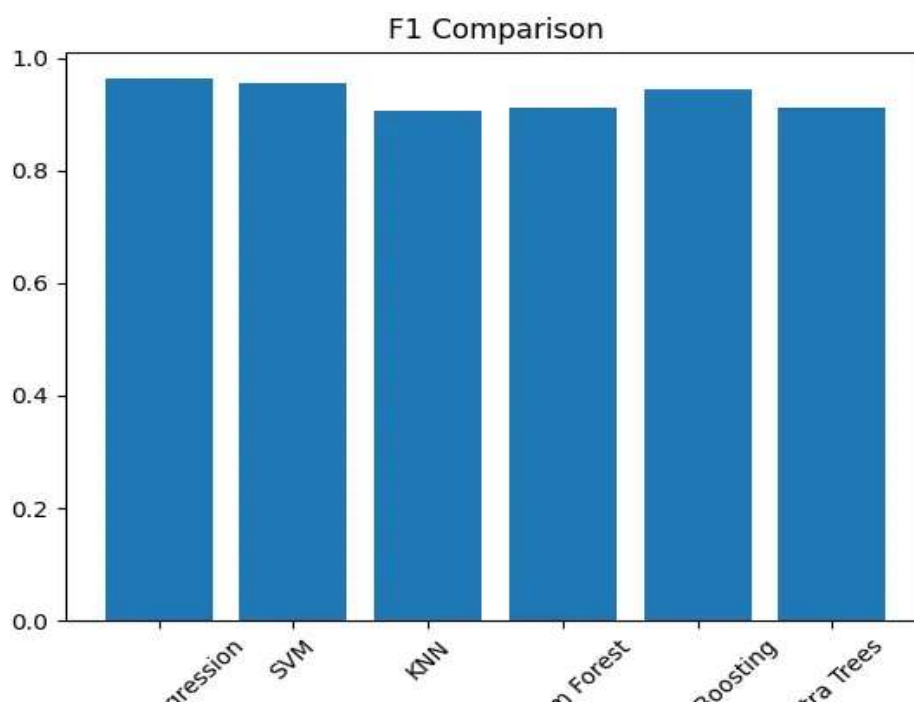


Fig.No.2: F1 Comparison

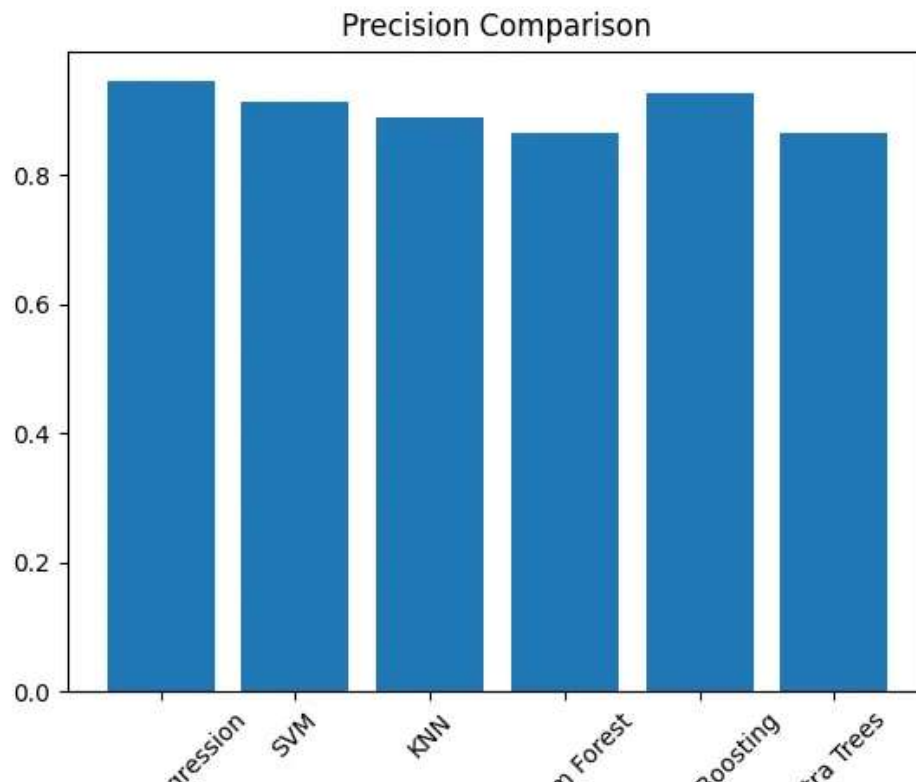


Fig. No.3:precision Coparison

V. DISCUSSION

The results generated from the Depression Prediction System confirm that machine learning models can indeed process student related behavioural, academic and psychological characteristics and thereby predict their risk for depression. Multiple classification algorithms were applied so as to arrive at the best suited model for the available data. The ensemble learning models like Random Forest and Extra Trees Classifier proved superior as they are effective in handling the relationships among various attributes and are less prone to over fitting problem.

It is noted from the study that the most important factors among students influencing their mental health are Academic pressure, Sleep duration, Financial stress, Study satisfaction and suicidal thoughts. Those students who suffer from immense workload in academic sector, wrong sleep duration habits, improper life style and emotional stress were susceptible for the depression risk signs. So it is clear that a depression among students has to be considered as a result of many interrelated factors not only a single factor.

Preprocessing methods applied in the system that includes One-Hot Encoding and Standard Scaling have contributed in better performance of the model and helped in enabling the machine learning models work in both categories and numerical inputs. The various evaluation metrics used such as Accuracy, Precision, Recall and F1-Score have given us more in-depth overview on the prediction power of each models.

The Gradio based application added to the user-friendliness of the system making it easy for the user to interact with the machine learning models. User can input data and compare the performance and output of various models with their visual results and also predict the risk level individually by inputting values. With this implementation a wider group of audience including those who have no technical background can now access the system.

Though the proposed system provides accurate prediction of the depression risk level, it has some limitations. Accuracy depends largely on the data available for training and also it depends on user inputs for predictions. Also the system is

not intended to be an alternative for a medical checkup or counseling and can serve as an early prediction tool only. Improvement to be considered in the system include having a larger dataset, using deep learning models, having a live tracking system, a mobile application support and using smart sensors and wearable devices for collecting data continuously.

VI.CONCLUSION

goal of this research work is to build a ML-based Depression Prediction System that examines the mental health of student and helps predict depression risk at an early stage. This system is developed by extracting some vital academic, behavior, psychological, and life-style based features like Academic pressure, Sleep duration, Study satisfaction, Financial stress, Dietary behavior, Suicidal thoughts, and Family history of mental illness to predict depression among students. The various ML algorithms like Logistic Regression, SVM, KNN, Random Forest, Gradient Boosting and Extra Trees Classifier were applied and tested. In the trained models, ensemble learning algorithms produced better prediction accuracy and performance. Data quality of the given features was improved by applying some preprocessing techniques.

The GUI which was built using Gradio has make the system much interactive and convenient. Through this UI a user can load the dataset, train the model, visualize the result and make prediction manually. Our system can identify whether a student have Low, Medium and High depression risk level and gives better visualization to user about student's mental health. Our study concluded that ML can be an efficient approach to the depression risk detection and analysis for mental health at university level.

Although the prediction is no way substitute the medical and clinical depression diagnosis and treatment, it can act as a support screening tool for prevention and intervention programs. More possible improvements to this system can be deep learning model, deployment on cloud and web, real-time monitoring, mobile application implementation, and expansion using big data.

REFERENCE

- [1] World Health Organization, *Depression and Mental Health Report*, Geneva, Switzerland: WHO, 2023.
- [2] J. Smith and P. Kumar, "Machine Learning Techniques for Mental Health Prediction," *International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 45–58, 2022.
- [3] T. Brown, A. Wilson, and R. Lee, "Student Depression Analysis Using Machine Learning Algorithms," *IEEE Access*, vol. 10, pp. 11234–11245, 2022.
- [4] H. Lee and S. Park, "Predictive Modeling of Depression Using Behavioral and Psychological Features," *Journal of Healthcare Informatics Research*, vol. 5, no. 2, pp. 89–102, 2021.
- [5] R. Johnson, "Impact of Sleep and Academic Stress on Student Mental Health," *Journal of Educational Psychology*, vol. 18, no. 1, pp. 21–35, 2020.
- [6] M. Patel and A. Sharma, "Ensemble Machine Learning Methods for Depression Prediction," in *Proc. International Conference on Data Science and Artificial Intelligence*, 2023, pp. 155–160.
- [7] Y. Zhang, "Explainable AI for Mental Health Prediction Systems," *Springer Journal of AI Applications*, vol. 8, no. 4, pp. 211–225, 2022.
- [8] S. Gupta and N. Rao, "Classification Techniques for Student Mental Health Prediction," *International Journal of Computer Applications*, vol. 176, no. 12, pp. 10–18, 2021.

- [9] K. Anderson and L. Martin, “Data Preprocessing Methods in Healthcare Machine Learning Systems,” *Journal of Data Science and Analytics*, vol. 14, no. 2, pp. 95–108, 2020.
- [10] P. Verma and S. Karthik, “Depression Risk Detection Using Ensemble Learning Approaches,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 233–241, 2022.