Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
June 2024 5(6), 2002–2009

COMPARATIVE STUDY OF MULTINOMIAL LOGISTIC REGRESSION AND RANDOM FOREST ALGORITHMS FOR PREDICTING PSYCHOLOGICAL WELLNESS AMONG STUDENTS' MENTAL HEALTH SURVEY

Pooja Mishra [1], Sunita Kushwaha [2]

[1] Research Scholar, Mats School of Information Technology, Mats University, Raipur (C.G.), India
[2] Assistant Professor, Mats School of Information Technology, Mats University, Raipur (C.G.), India

Corresponding Author
Pooja Mishra,
shubhpooja1919@gmail.com

DOI
10.29121/shodhkosh.v5.i6.2024.5143

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Stress represents a common emotional reaction to stimulating conditions however, when it becomes excessive particularly among students, it may result in serious mental health syndromes, physical uneasiness, and self-harm. This research aims to classify student wellness levels by inspecting a broad range of stress-related proportions, including psychological, physiological, academic, social, and ecological aspects. This study investigates student psychological wellness using machine learning techniques. A mental health survey dataset from Kaggle was analysed to predict wellness levels based on stress, anxiety, depression, sleep, diet, physical activity, and social factors. A Multinomial Logistic Regression (MLR) model and a Random Forest (RF) model were applied to the data. A Wellness Level was calculated to classify students into Low, Average, Moderate, and High wellness categories. Performance of both models was evaluated using accuracy, precision, recall, F1-score, and AUC-ROC. Results indicated that the MLR model achieved slightly higher predictive accuracy than RF. The findings suggest that early mental health screening can be effectively supported by predictive models. This contributes to proactive student care through targeted interventions. The models help institutions allocate resources more efficiently. Overall, machine learning presents a valuable approach for psychological wellness prediction. The study promotes data-driven decisions in academic mental health strategies.

Keywords: Psychological Wellness, Machine Learning, Multinomial Logistic Regression, Multinomial Random Forest, Mental Health Survey, Classification

1. INTRODUCTION

Psychological strength has occurred as a serious concern for students, with growing academic burdens, routine variations, and socioecological challenges meaningfully affecting their emotional comfort. Stress, anxiety, depression, and poor sleep or dietary habits are gradually reported among young beginners, affecting not only their academic performance but also their overall eminence of life. The initial recognition of mental health matters and the enactment of effective intervention approaches have become essential for helping student wellness.

The atmosphere during school and college is positively very diverse. College students will find different learning methods compared to their school days. Mental health for college students who are consecutively the lecture process is imperative in order to complete their academic tasks well. For final year students who are in the progression of studying, proposition is one of the reasons of psychological health problems [1,2].

Recent developments in data science and machine learning have unlocked new possibilities for diagnostic exhibiting in the field of mental health. machine learning algorithms helps in the context of depression prediction, utilizing the growing availability of mental health data to evaluate it very well [3]. These tools proposition an organized and mountable approach to identify patterns in complex datasets, allowing institutions to recognize early cautionary signs of psychological distress. the classification of student mental health data, where mental health data is used as input for a model to develop and apply for classifying data to evaluate outcome to find that student comes under which category [4]. By leveraging survey data on various wellness pointers such as stress levels, anxiety, depression, sleep quality, physical activity, diet, and social support, machine learning models can classify students into separate wellness categories.

2. LITERATURE REVIEW

Akash et al. this study is to compare the efficacy of random forest and logistic regression in order to enhance stress level analysis through the use of supervised learning techniques. While random forest obtained an accuracy of 85.236%, logistic regression was able to attain 94.448%. The results demonstrate that because logistic regression has a higher accuracy than random forest, it performs better in identifying fraudulent services. [5].

Dong The ML algorithms in this study included random forest (RF), support vector machine (SVM), neural network (NN), and gradient boosting machine (GBM) methods. LR and NN had the best performance in terms of AUCs. Compared with ML models, LR model performed comparably to ML models in predicting depressive symptoms and identifying potential risk factors while also exhibiting a lower risk of overfitting [6].

Deena et al. study examined that Several machine learning algorithms Support Vector Machines, Logistic Regression, Naive Bayes, Decision Trees, and Random Forest are examined, with Naive Bayes identified as the most effective model. It attains a prediction accuracy of 90% [7].

Saxena et al. showed that the Naïve Bayes classifier, without feature selection, achieved the highest TPR (True Positive Rate) of 0.923, meaning it accurately identified most cases of depression. the Logistic Regression model demonstrated promising results, particularly when feature selection techniques were employed [8].

Sahlan et al. This study explores the state of student well-being and uses various machine learning algorithms for the prediction process based on the data of entrepreneurial competency in university students. The results indicate the choice of major and gender has a significant impact on a student's well-being. Decision Tress perform better than KNN and SVM presenting an accuracy [9].

Yang et al. This study explores the results of extensive performance evaluation for the proposed ES-ANN model are performed by applying well-known G-mean and F1 score performance metrics. The results indicated that the ES-ANN model outperformed state-of-the-art benchmark methods, namely Random Forest and Decision Tree [10].

Luo et al. study that RFA identified several factors associated with depression risk, with suicidal ideation, anxiety, and sleep quality exhibiting the strongest associations. Other significant predictors included academic stress, BMI, vital capacity, psychological resilience, physical fitness test scores, major satisfaction, and social network use. The model achieved an accuracy of 87.5% [11].

Basysyar et al. studied that the impact of emotions and mental health on students' The results indicated that the Logistic Regression model achieved an accuracy of 86.55%, while the Random Forest model achieved a slightly higher accuracy of 87.62% [12].

3. METHODOLOGY

The primary objective of this study is to compare the performance of Multinomial Logistic Regression (MLR) and Random Forest (RF) algorithms in predicting the psychological wellness of students based on a mental health survey dataset. The methodology involves the following steps:

- **Data Extraction:** Selected only relevant features using feature engineering to optimize memory usage.
- **Data Storage:** Collected data stored in Excel, converted to CSV for processing in R.
- **Tool Used:** R programming was used for developing and executing sentiment prediction models.

- **Model Implementation:**
  - Applied Multinomial Logistic Regression and Random Forest models.
  - Split data into training and testing sets.
- **Prediction:** Predicted psychological wellness levels (High, Moderate, Average, Low) for each instance.
- **Evaluation:** Compared predicted vs. actual values.
- **Validation:** Assessed model performance using accuracy, precision, recall, and F1-score.

Figure 1

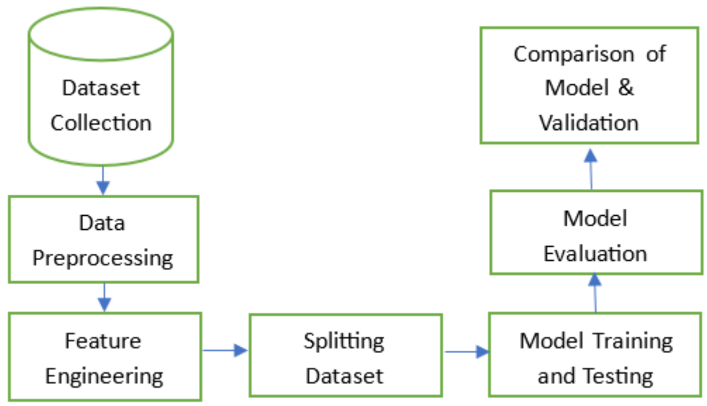

Figure 1 Design Workflow

3.1. DATASET

The dataset utilized in this study was sourced from Kaggle and contains a total of 7,022 student records related to mental health indicators. These data points capture various psychological and lifestyle-related parameters including stress, anxiety, depression, sleep quality, diet, physical activity, and social support. After preprocessing, the dataset was split into two parts: 4,948 records 70% were used for training, and 2,074 records 30% were reserved for testing. For the purpose of multinomial classification, the target variable Wellness Level was categorized into four distinct classes are High Wellness (5,819 records), Moderate Wellness (639 records), Average Wellness (75 records), Low Wellness (9 records). here Two machine learning algorithms Multinomial Logistic Regression (MLR) and Random Forest (RF) were applied to classify the students into one of the wellness categories. These classifications help in predicting the psychological state of students with respect to multiple influencing factors.

3.2. MODEL DEVELOPMENT

1) **Multinomial Logistic Regression (MLR):** The MLR model was developed to predict the psychological wellness categories. The model was trained using the selected features, with the target variable being the psychological wellness classification.

MLR models the log-odds for each class relative to the reference category. For a categorical target variable Y with K categories (e.g., high, moderate, average, low), the model would compute the log-odds of being in each category relative to the reference class.

$\log(P(Y=k) / P(Y=low)) = \beta 0k + \beta 1\ X1 + \beta 2\ X2 + \cdots + \beta p\ Xp$ for $k=1, 2, \ldots, K-1$

Here, $P(Y=k)$ is the probability of the observation belonging to category k (e.g., High, Moderate, Average).

$P(Y=Low)$ is the probability of the observation belonging to the reference category (Low).

The coefficients βik represent the relationship between the independent variables ($X1, X2, \ldots, Xp$) and the log-odds of being in category k relative to the reference [13,14].

2) **Random Forest (RF):** The Random Forest is ensemble classifier using many decision tree models. Random forest combines the idea of bootstrapping data from a learning dataset to form training data set and selecting

parameters randomly to construct decision trees [15]. In Random Forest classification, the algorithm aggregates predictions from multiple decision trees using a majority vote.

Let's say we have a Random Forest classifier consisting of K decision trees. Each decision tree Tk predicts the class label for a given in put sample x as $\hat{y}$ k, where k=1,2,…,k=1,2,…,K.

The Random Forest prediction y̑k for the input sample x is determined by the majority vote among the predictions of all decision trees:

$$\hat{y}=\text{mode}\ (\hat{y}1,\hat{y}2,\ldots\ldots\ldots,\hat{y}k)$$

Were,

- y̑i is the predicted class label by the ith decision tree.
- Mode represents the function that returns the most common label model among the predictions.

Majority voting scheme helps to ensure robustness and improve the generalization performance of the Random Forest classifier.

3.3. MODEL EVALUATION

Both models were evaluated using a set of performance metrics, including accuracy, precision, recall, F1 score, and the Area Under the Receiver Operating Characteristic (AUC-ROC) curve. These metrics were computed on the test set to compare the models' ability to classify psychological wellness levels accurately.

3.4. IMPLEMENTATION TOOLS

All models were implemented using R Language using its required libraries and packages for machine learning algorithms like dplyr, pROC, ROCR, randomForest etc. for data manipulation, and caret for data visualization. The models were trained and evaluated on a machine with sufficient computational resources to handle the dataset size.

4. RESULTS AND DISCUSSION

After training both models, the results will be compared, and insights into which algorithm provides better prediction accuracy for psychological wellness will be discussed. The effectiveness of each algorithm will be analysed in terms of interpretability, robustness, and scalability. Both models were evaluated using standard classification metrics: Accuracy, Precision, Recall, F1-score and Area Under the Curve (AUC) in Table 1 shows the results using evaluation metrics.

Table 1

Results of Evaluated Matrices

Model	Accuracy	Precision	Recall	F1- Score	AUC
Logistic Regression	97.7%	96.6%	80.0%	81.5%	86.7%
Random Forest	96.4%	87.5%	53.8%	60.1%	84.7%

Multinomial Logistic Regression outperforms Random Forest across all evaluation metrics. It also demonstrated excellent precision (96.60%), recall (80.00%), and F1-score (81.50%), suggesting a strong balance between identifying true positives and minimizing false classifications. The AUC score (86.70%) confirms Logistic Regression's better discriminatory power in distinguishing between different psychological wellness levels.

Figure 2

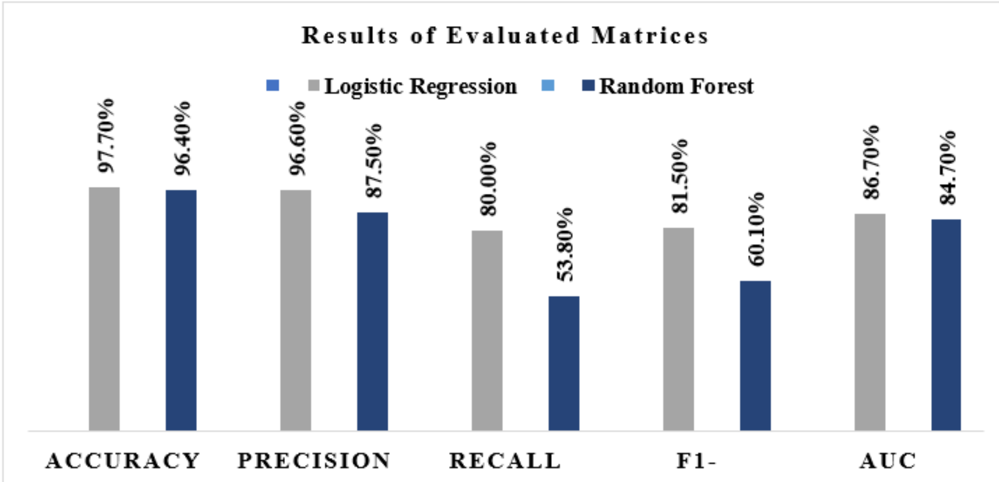

Figure 2 Graph of Evaluated Matrices

Figure 3

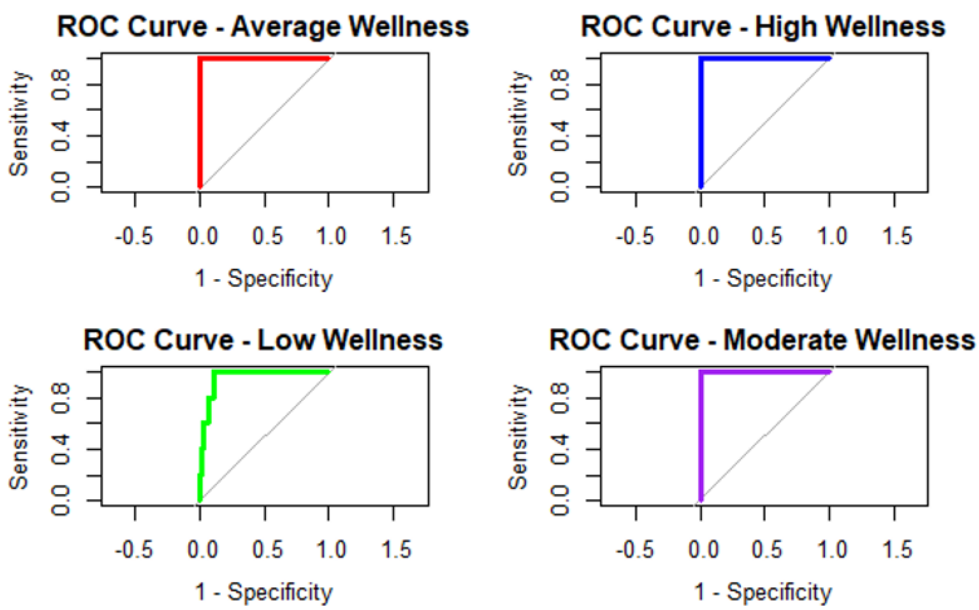

Figure 3 Logistic Regression Model Levels of wellness (Average, High, Low, Moderate)

Table 2

Logistic Regression Model Levels of Wellness AUC Percentage		
No.	**Classes**	**AUC Percentage**
1	Average Wellness	99.79
2	High Wellness	100.0
3	Low Wellness	95.82
4	Moderate Wellness	100.0

MLR reached perfect AUCs (100%) for High and Moderate Wellness, but performed lower on Low Wellness, indicating slight variability in prediction accuracy.

Figure 4

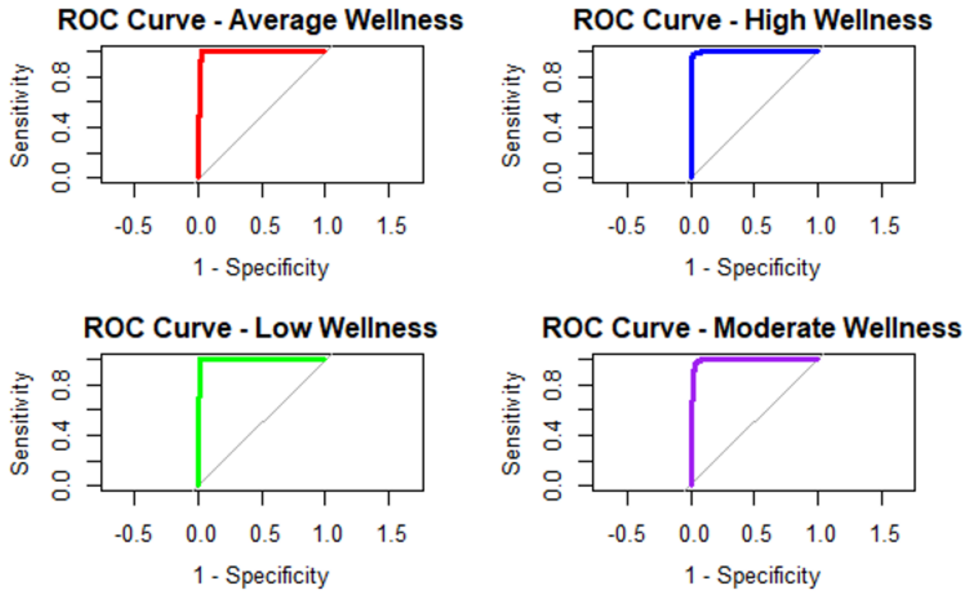

Figure 4 Random Forest Model Levels of wellness (Average, High, Low, Moderate)

Table 3

Random Forest Model Levels of Wellness AUC Percentage		
No.	**Class**	**AUC Percentage**
1	Average Wellness	99.13
2	High Wellness	99.72
3	Low Wellness	99.45
4	Moderate Wellness	98.8

Whereas Random Forest model achieved higher AUCs for Low Wellness (99.45% vs. 95.82%) and showed more balanced performance across all classes.

After comparing both models conclude that Multinomial Logistic Regression is a more reliable and accurate model for predicting psychological wellness among students in the context of the given dataset. This makes it a suitable choice for practical deployment in educational or mental health settings aiming to classify and monitor students' mental well-being.

5. CONCLUSION

This study conducted a comparative analysis of two supervised machine learning algorithms Multinomial Logistic Regression (MLR) and Random Forest (RF) to predict psychological wellness levels among students based on mental health survey data. The primary objective was to classify students into one of four Wellness categories: High, Moderate, Average, and Low psychological wellness. The models were evaluated using key performance metrics: Accuracy, Precision, Recall, F1-Score, and AUC. The findings clearly indicate that Multinomial Logistic Regression showed superior accuracy 97.7% constantly outperforms as compared to Random Forest across all evaluated metrics. These results suggest that MLR is more suitable for this multi-class psychological wellness classification task, especially in contexts where initial identification of mental health concerns is dangerous. Multinomial Logistic Regression offers a more effective and interpretable solution for predicting students' psychological wellness and can be considered for integration into student support systems for mental health monitoring and early intervention strategies.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Wijaya, Vannes & Rachmat, Nur. (2024). Comparison of SVM, Random Forest, and Logistic Regression Performance n Student Mental Health Screening. JEECS (Journal of Electrical Engineering and Computer Sciences). 9. 173-184. 10.54732/jeecs.v9i2.9.

M. K. Sari and E. A. Susmiatin, (2023), "Deteksi Dini Kesehatan Mental Emosional pada Mahasiswa," *Jurnal Ilmiah STIKES Yarsi Mataram*, vol. 13, no. 1, pp. 10–17, doi:10.57267/jisym.v13i1.226

M. Rijal, F. Aziz, and S. Abasa, (2024), "Prediksi Depresi : Inovasi Terkini Dalam Kesehatan Mental Melalui Metode Machine Learning Depression Prediction : Recent Innovations in Mental Health Journal Pharmacy and Application," *Journal Pharmacy and Application of Computer Sciences*, vol. 2, no. 1, pp. 9–14.

H. D. Putra, L. Khairani, and D. Hastari, (2023), "Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa," in *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, pp. 120–125

Akash, T. & Arul, U. (2025). Effective analysis of stress level using logistic regression compared with random forest. 10.1201/9781003559115-81.

Xing-Xuan Dong1, Jian-Hua Liu1, Tian-Yang Zhang1,2,3, Chen-Wei Pan1, Chun-Hua Zhao4 ✉, Yi-Bo Wu5, and Dan-Dan Chen6,7," Comparison of Logistic Regression and Machine Learning Approaches in Predicting Depressive Symptoms: A National-Based Study", ISSN 1976-3026

G. DEENA[1]*, A. SANDHYA2, K. RAJA3(2024)," Machine Learning-Based Classification and Prediction of Student Stress Levels: A Comparative Study of Algorithms", Journal of Theoretical and Applied Information Technology, ISSN: 1992-8645, Vol.102. No. 19

Rajesh Saxena[1], Dr. Ashish Saini2," Comparative Analysis of Machine Learning Strategies for Depression Detection in Indian College Students ", Nanotechnology Perceptions 20 No.7 (2024) 617–633

Fadhluddin Sahlan[1], Faris Hamidi[2], Muhammad Zulhafizal Misrat3, Muhammad Haziq Adli4, Sharyar Wani5, Yonis Gulzar6," Prediction of Mental Health Among University Students", International Journal on Perceptive and Cognitive Computing (IJPCC), Vol 7, Issue 1 (2021)

Liangqun Yang[1], Haibin Ni[2], Yingdong Zhu*[3]," Data-Driven Mental Health Assessment of College Students Using ES-ANN and LOF Algorithms During Public Health Events ", (2025) 59–76

Lin Luo[1]*, Junfeng Yuan[1], Chenghan Wu[2], Yanling Wang1, Rui Zhu1, Huilin Xu1, Luqin Zhang1 and Zhongge Zhang1," Predictors of depression among Chinese college students: a machine learning approach", Luo et al. BMC Public Health (2025)

F. M. Basysyar, G. Dwilestari, and A. I. Purnamasari, (2024), "Analysis Student Emotions And Mental Health on Cumulative GPA Using Machine Learning and Smote," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 2, pp. 361–368,

Agresti, A. (2002). *Categorical Data Analysis*. Wiley-Interscience.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley.

mohammed nazim uddin, md.ferdous bin hafiz2 sohrabhossain shah mohammadmominulislam, drug sentiment analysis using machine learning classifiers, (ijacsa) international journal of advanced computer science and applications, vol. 13, no. 1,2022

Ujunwa Madububambachu[1],* , Augustine Ukpebor[2] and Urenna Ihezue[3],"Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review",(2024), ISSN: 1745-0179

Anbarasan, Padmavathi and Rohith, P. Sreeram and Pranav, R. and Surya, A.V.K Sai, Mental Health State Identification using Machine Learning Algorithms (March 27, 2025). Available at SSRN: 5195594

Algorithms Dr. S. Bharathidason[1] and C. Sujdha[2], "A Comparative Study of Mental Health Prediction using Machine Learning", 2024 JETIR March 2024, Volume 11, Issue 3

Khurana, Yashika and Jindal, Sachin and Gunwant, Harsh and Gupta, Deepak, Mental Health Prognosis Using Machine Learning (March 17, 2022). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2022, Available at SSRN: 4060009

Bernal-Salcedoc, J., Vélez Álvarez, C., Tabares Tabares, M. *et al.* Classification of depression in young people with artificial intelligence models integrating socio-demographic and clinical factors. *Curr Psychol* (2025).

A. Tabassum, Y. F. Ema, M. M. S. Rafee and M. S. I. K. Limon, "A Survey on Predicting Depression Among Teachers and Students at Metropolitan University (Sylhet, Bangladesh) Using Machine Learning," *2025 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, Bangalore, India, 2025, pp. 1-6, doi: 10.1109/IITCEE64140.2025.10915517.