ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research

Vol. 17, Issue, 01, pp.31261-31264, January, 2025
DOI: https://doi.org/10.24941/ijcr.48311.01.2025

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

SCHOLARLY ARTICLES FOR COMPARISON OF XG BOOST CLASSIFIER WITH LOGISTIC REGRESSION ALGORITHM TO IMPROVE ACCURACY IN STUDENT MENTAL HEALTH PREDICTION

[1]Sakthivel S and [2,*]Raja, S. R.

[1]Research Scholar, Department of Computer Applications, Centre for Open and Digital Education, Hindustan Institute of Technology and Science, India; [2]Associate Professor, Master of Computer Applications, Center for Open and Digital Education, Hindustan Institute of Technology and Science, Chennai, India

ARTICLE INFO

Article History:
Received 20th October, 2024
Received in revised form
17th November, 2024
Accepted 24th December, 2024
Published online 30th January, 2025

Key Words:

Health, Prediction, Logistic Regression, Xgboost Classifier, Machine Learning, Algorithm, Accuracy.

***Corresponding author:* Raja, S. R.**

ABSTRACT

Aim: The purpose of this study is to develop a reliable forecast model for student mental health. This model would offer important insights to educators and mental health professionals, helping them identify and support students who may be struggling with their mental health. By doing so, we hope to promote better mental health outcomes for students. **Materials and Methods:** This study assesses the predictive capability of XGBoost Classifier andLogistic Regression models for student mental health, using the "Student Mental Health" dataset obtained from Kaggle. The dataset comprises 10 pertinent columns of information. Prior to analysis, the dataset underwent several preprocessing steps including feature scaling, one-hot encoding, and outlier management, and was subsequently split into training and testing sets. The models' predictive performance was assessed through accuracy metrics and statistical analysis was conducted utilizing SPSS software. The sample sizes for both groups were calculated using clincalc.com. **Results:** This findings of this study indicates that machine learning algorithms are capable of accurately predicting a Mental Health of a Student. The significance value for this study is $p=0.001$, where is $p<0.05$. Hence, there is a statistically substantial dissimilarity between the two groups. **Conclusions:** In conclusion, findings from this investigation indicate that the Logistic Regression model has exhibited promise and efficacy in precisely predicting student mental health. Therefore, it could be advantageous to delve deeper into this approach in future studies.

Copyright©2025, Sakthivel S and Raja. *This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

Citation: Sakthivel S and Raja, S. R. 2025. "Scholarly articles for comparison of XG Boost Classifier with Logistic Regression Algorithm to Improve Accuracy in Student Mental Health Prediction*". International Journal of Current Research*, 17, (01), 31261-31264.

INTRODUCTION

A mental illness, also known as a mental disorder, refers to a health condition that alters an individual's thinking, feelings, or behavior, causing them distress and making it challenging to lead a regular life (Fadhluddin Sahlan1, Faris Hamidi 2, Muhammad Zulhafizal Misrat3, Muhammad Haziq Adli4, Sharyar Wani5 , Yonis Gulzar, n.d.). University students face a heightened risk of experiencing elevated stress levels due to the rigorous academic demands coupled with limited time for social and personal activities (Mello 2016). This can exacerbate the usual stressors of life and have an adverse impact on their overall well-being(Mohd and Yahya 2018). The technique of machine learning employs sophisticated statistical and probabilistic techniques to create systems that enhance their performance through experience. It is widely regarded as a valuable tool in the prediction of mental health, enabling researchers to extract essential insights from data, customize experiences, and build intelligent automated systems (Chung and Teo 2022). The purpose of this paper is to explore the status about student well-being and employ numerous machine learning algorithms forecast mental fitness wellbeing of university students using relevant data (LAbate 2012). The subject of our research, which involves predicting Novel Student Mental Health through machine learning, has been extensively studied, with approximately 36 relevant studies on IEEE Xplore and around 8658 results on Science Direct over the past five years. We came across a multitude of studies during our research that were relevant to our inquiry and proved to be advantageous. Among them the most cited was(Chung and Teo 2022; Hasanbasic *et al*. 2019), they got a higher accuracy of 91% using SVM and this study used a relatively small dataset and equipment to measure stress and anxiety levels in students before exams and presentations. The outcomes disclosed, every student came across high levels of stress and anxiety. Pre-test and test conditions can damage students' physical and mental health was illustrated in the analysis. To build an automated stress detection system wearable sensors can be used. Another notable study which was most useful is (Mohd Shafiee *et al*. 2020) they collected algorithms and accuracies of multiple studies related to this study and compared them. Mental Health is a common problem that affects many people in their daily lives. In previous studies in prediction of Mental Health of a Student, accuracy of XG Boost is not optimal.

So, this study's objective is to identify Mental Health of a Student through the comparison of XGBoost and Logistic Regression in order to increase the precision.

MATERIALS AND METHODS

The current study consists of two groups, with group 1 using the XGBoost Classifier algorithm and group 2 using the Logistic Regression algorithm to predict Novel Student Mental Health. The study compared efficiency of two algorithms-Logistic Regression and XGBoost Classifier, which predicts Novel Student Mental Health using a dataset obtained from Kaggle. The data set was obtained from a survey conducted among students in a university, and it included attributes such as their gender, age, course, year of study, CGPA, marital status, depression status, anxiety status, panic attack status, and whether they sought any specialist for treatment. These attributes were considered as independent variables for predicting the target variable, which was the overall mental health status of the students.

The dataset was obtained from the Kaggle Dataset Novel Student Mental Health dataset in and the sample size of 118 patient records was the same for both groups and the sample size for both algorithms was determined using pre-test power analysis with an 80% power and an alpha value of 0.05. The study aimed to find the algorithm that provides better accuracy in predicting Novel Student Mental Health. Group 1: The first group consisted of a sample of 10 records, which were preprocessed to remove any null or missing values and outliers that could have skewed the analysis. The data was then divided into two sets, with 70% of the sample used for model training, and the remaining 30% used to test the model's performance. By following this procedure, a fair evaluation of the model's accuracy and reliability could be obtained. Group 2: In the second group, a similar approach was adopted. A sample of 10 records was taken, and the data was preprocessed to eliminate any null or missing values and outliers. The parameters mean and standard deviation from the previous established paper were to be the input for sample size calculation (Saito, Suzuki, and Kishi 2022).

The sample was then split as two sets, where 70% used for model training and 30% reserved for model testing. This method ensured that the model was trained with representative data subset and its performance was accurately evaluated without bias. Overall, the same procedures were followed for both groups to ensure consistency and fairness in the analysis.
In this study, Google Colab was utilized as a cloud-based platform with ample resources such as 12GB of RAM and 107 GB of disk space. This eliminated the need for additional hardware resources. Data preprocessing was conducted at the start of the study to remove null or missing values and outliers from the dataset. The dataset was then split as two parts, where 70% used for model training and 30% reserved for performance testing. XGBoost Classifier algorithm was used for Group 1, while the Logistic Regression algorithm was used for Group 2. The models' accuracy of prediction, F1 Score, Precision, and Recall Score were evaluated using the test data. This same testing procedure was conducted for both groups, ensuring a fair performance evaluation of the two algorithms in forecasting heart disease in the patient dataset. The proposed work architecture is shown in Fig. 1.

Statistical Analysis: The proposed Logistic Regression algorithm and existing XGBoost Classifier were utilized to analyze the accuracy of Novel Student Mental Health prediction. To compare accuracy, mean, standard deviation, standard error of the two algorithms t-tests were computed using SPSS software version 29. An independent-sample-t-test was conducted to define the existence of significance among two algorithms with a 95% confidence interval.

RESULTS

Table 1 Independent samples comparing XGBoost Classifier for student mental health prediction with Logistic Regression. In XG Boost for predicting student mental health, the mean accuracy is 66.67%, whereas in Logistic Regression it is 79.52%. XG Boost for predicting student mental health has a standard deviation of 6.35083 and Logistic Regression has standard deviation of 2.00831

	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	XGBoost Classifier	10	66.6670	6.35083	2.00831
	Logistic Regression	10	79.5230	7.11522	2.25003

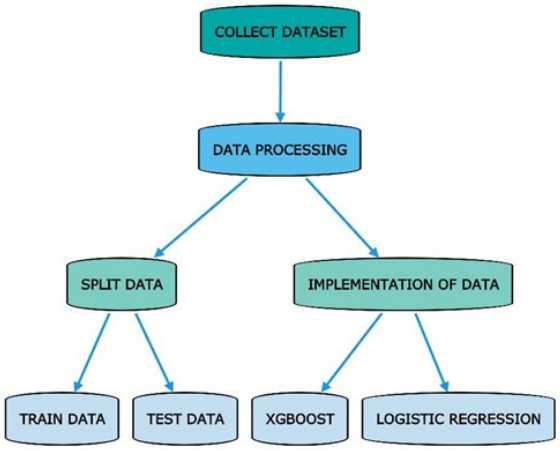

Fig. 1. Flow chart depicting the methodology adopted in the study

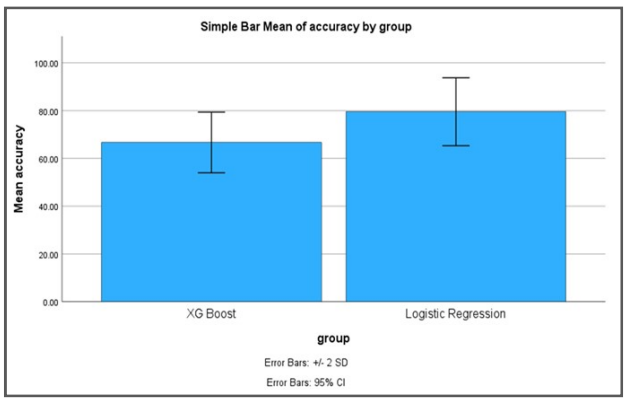

Fig. 2. Bar plot showing the mean accuracy plotted for the two groups considered, XGBoost Classifier andLogistic Regression. The mean accuracy is better for the logistic Regression classification than the XGBoost Classifier Model. X-axis: XGBoost Classifier vsLogistic Regression (two groups) and Y-axis: Mean accuracy of classification with error bars of ± 2 SD, Error Bars: 95% CI

Table 2. Independent samples for detecting Students mental health with Logistic Regression algorithm, 95% confidence interval 0.001 (p>0.05) significance value, So, there is a statistically significant difference between the two groups

		Levene's Equality of variances		T-test for equality of means					95% Confidence interval of the difference	
						Significance				
		F	Sig	t	df	One-sided p	Two-sided p	Mean Difference	Std. Error Difference	Lower	Upper
Accuracy	Equal variances assumed	0.180	0.676	-4.263	18	<0.001	<0.001	-12.85600	3.01595	-19.19227	6.51973
	Equal variances not assumed			-4.263	17.7772	<0.001	<0.001	-12.85600	3.01595	-19.19809	6.51391

Table 3. Classification report between XG Boost Classifier and Logistic Regression

Algorithm	Accuracy	F1 Score	Precision	Recall
XGBoost Classifier	66.67%	0.6615	0.5606	0.5474
Logistic Regression	79.52%	0.7760	0.9083	0.5082

DISCUSSION

The results showed that the proposed model of Logistic Regression outperformed the existing XGBoost Classifier model in predicting student mental health. The Logistic Regression classification model reached an accuracy of 79.52%, where the XGBoost Classifier model achieved 66.667%. It has been observed that the proposed model of Logistic Regression outperformed the existing model of XG Boost in predicting student mental health. (Mohd Shafiee *et al*. 2020; Sofianita Mutalib1, Nor Safika Mohd Shafiee2, Shuzlina Abdul-Rahman, n.d.) have reported that Logistic Regression has achieved a remarkable accuracy of 87.41% , whereas (Bunting *et al*. 2023), (Jang and Kim 2023), (Foster *et al*. 2023), (Wong *et al*. 2023) also reported that Logistic Regression has performed well , while comparing with other algorithms which are similar to our study.(Du 2022) have reported that XG Boost resulted in an accuracy score of 52% which is significantly lower but has performed better than several algorithms such as KNN, SVM and Random Forest.(Vaishnavi *et al*. 2022) have reported that Logistic Regression with an accuracy of 79% did not perform as well as other algorithms such as KNN and Random Forest with only around 1% accuracy difference which is against our findings.

Despite its limitations, this study's findings hold considerable promise in forecasting student mental health. The outcomes underscore the significance of selecting the most fitting algorithm for predicting student mental health, taking into account the dataset's characteristics and the research question under investigation. In general, this study emphasizes the crucial role of meticulous algorithm selection in data analysis to guarantee precise predictions. Our research focuses on forecasting student mental health, a crucial field of study in medicine. Various factors, including environmental influences, dataset quality, preprocessing methods, and algorithm choice, may account for the inconsistencies in these findings. The accuracy of the predictions is heavily influenced by the dataset's quality and representativeness. Moreover, the selection of preprocessing techniques, such as handling null or missing values and outliers, can have a considerable impact on the results. Ultimately, choosing the appropriate algorithm must be thoughtfully considered based on the dataset's characteristics and the research question being addressed. Future studies can therefore concentrate on improving these elements to raise the precision of predicting a student's mental health of a student.

CONCLUSION

In Conclusion, the study suggests that theLogistic Regression model outperformed XG Boost Classifiers in forecasting student mental health. This outcome underscores the potential effectiveness ofLogistic Regression as a valuable method for predicting student mental health and may warrant additional investigation in future research.

REFERENCES

Bunting, Lisa, Claire McCartan, Gavin Davidson, Anne Grant, Ciaran Mulholland, Dirk Schubotz, Ryan Hamill, *et al*. 2023. "The Influence of Adverse and Positive Childhood Experiences on Young People's Mental Health and Experiences of Self-Harm and Suicidal Ideation." *Child Abuse & Neglect* 140 (April): 106159.

Chung, Jetli, and Jason Teo. 2022. "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges." *Applied Computational Intelligence and Soft Computing* 2022 (January): 1–19.

Du, Wei. 2022. "Application of Improved SMOTE and XGBoost Algorithm in the Analysis of Psychological Stress Test for College Students." *Journal of Electrical and Computer Engineering* 2022 (May): 1–8.

Fadhluddin Sahlan1 , Faris Hamidi2 , Muhammad Zulhafizal Misrat3 , Muhammad Haziq Adli4 , Sharyar Wani5 , Yonis Gulzar. n.d. "Prediction of Mental Health Among University Students." https://journals.iium.edu.my/ kict/ index. php/ IJPCC/article/view/225/140.

Foster, Simon, Natalia Estévez-Lamorte, Susanne Walitza, and Meichun Mohler-Kuo. 2023. "The Impact of the COVID-19 Pandemic on Young Adults' Mental Health in Switzerland: A Longitudinal Cohort Study from 2018 to 2021." *International Journal of Environmental Research and Public Health* 20 (3). https://doi.org/ 10.3390/ ijerph20032598.

Hasanbasic, Amir, Mustafa Spahic, Dino Bosnjic, Haris H. Adzic, Vedad Mesic, and Omar Jahic. 2019. "Recognition of Stress Levels among Students with Wearable Sensors."

In *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*. IEEE. https://doi.org/ 10.1109/infoteh.2019.8717754.

Jang, Sou Hyun, and Juyeon Kim. 2023. "Stress or Buffer: The Impact of Social Transnational Ties on Depressive Mood and Suicidal Ideation Among Female Marriage Migrants in South Korea." *Journal of Immigrant and Minority Health / Center for Minority Public Health*, February. https://doi.org/10.1007/s10903-023-01457-6.

LAbate, Luciano. 2012. *Mental Illnesses: Understanding, Prediction and Control*. BoD – Books on Demand.

Mello, Joshua. 2016. *Life Adversity, Social Support, Resilience, and College Student Mental Health*.

Mohd, Norhatta, and Yasmin Yahya. 2018. "A Data Mining Approach for Prediction of Students' Depression Using Logistic Regression and Artificial Neural Network." In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*. New York, NY, USA: ACM. https://doi.org/ 10.1145/3164541.3164604.

Mohd Shafiee, Nor Safika, Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia, Sofianita Mutalib, and Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia. 2020. "Prediction of Mental Health Problems among Higher Education Student Using Machine Learning." *International Journal of Education and Management Engineering* 10 (6): 1–9.

Saito, Tomoki, Hikaru Suzuki, and Akifumi Kishi. 2022. "Predictive Modeling of Mental Illness Onset Using Wearable Devices and Medical Examination Data: Machine Learning Approach." *Frontiers in Digital Health* 4 (April): 861808.

Sofianita Mutalib1 , Nor Safika Mohd Shafiee2 , Shuzlina Abdul-Rahman. n.d. "Mental Health Prediction Models Using Machine Learning in Higher Education Institution." *TURCOMAT*. https://turcomat.org/index.php/ turkbilmat/ article/ view/2181

Vaishnavi, Konda, U. Nikhitha Kamath, B. Ashwath Rao, and N. V. Subba Reddy. 2022. "Predicting Mental Health Illness Using Machine Learning Algorithms." *Journal of Physics. Conference Series* 2161 (1): 012021.

Wong, Shun Sun, Charng Choon Wong, Kwok Wen Ng, Mohammad F. Bostanudin, and Suk Fei Tan. 2023. "Depression, Anxiety, and Stress among University Students in Selangor, Malaysia during COVID-19 Pandemics and Their Associated Factors." *PloS One* 18 (1): e0280680.

****