Available online @ https://jjem.jnnce.ac.in
https: www.doi.org/10.37314/JJEM.SP0257
Indexed in International Scientific Indiexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 08 December 2024

Student Mental Health Assessment via Machine Learning Techniques

Srinidhi R.K[1], Adarsh M.J[2],

[1]Student, [2]Assistant Professor, Department of Computer Application
JNN College of Engineering Shivamogga

rksrindhi@gmail.com, adarshmj@jnnce.ac.in

Abstract

The consequences of anxiety, stress and other pressures of the modern world affect people on the international level. In a recent study, it was established that technology in health care was capable of identifying much more data concerning the human body than the conventional measurement methods. Due to the large data set in the healthcare industry, there is no better approach to passing through such data than using an ML algorithm. To be more precise, an innovative approach of artificial intelligence's application in the sphere of mental health is to estimate the probability of particular diseases and to offer some recommendations for action. Huge and detailed this information is, and therefore machine learning has become one of the primary tools for healthcare analysis. issues and relationships in data that manual analysis often cannot discover are areas where the broad functions of ML prove useful.

The pressure in the contemporary society, worry, and stress on a worldwide level. A similar study was published revealing that hi-tech in health care can obtain significantly rich and denser data about human physical form as those obtained by conventional measurement. Based on the data volume in health care, it is possible to identify that only machine learning (ML) algorithms are suitable for data processing. In detail, the use of artificial intelligence (AI) in mental health care has been more revolutionary since they determine the likelihood of such diseases and offer subsequent suggestions. The tremendous amount of data and its specificity and comprehensiveness require the use of machine learning algorithms, as they perfectly indicate problems and relationships in large datasets that cannot be identified by ordinary examination. Out of all the ML methods, the random forest, decision tree, and logistic regression affairs significantly for healthcare investigation improving the efficacy of estimates and suggestions in mental health.

Keywords: *Machine learning, random-forest, decision-tree, logistic regression*

1. Introduction

It is needed to examine behavioural indications such as relationships, academic achievement, and interactions with the environment with respect to identify depression in kids. Abrupt behavioural changes, chronic fatigue, trouble concentration, and extended absence are indicators of depression. Teachers and mental health specialists need to keep a close eye on these indicators with respect to recognize and

assist impacted pupils. Research highlights the utilization of behavioural data to create prediction models for early depression identification, including attendance, academic performance, and social interactions. By taking a proactive stance, prompt intervention is made possible, improving student wellbeing and creating a positive school climate. Important Indicators of depression, such as tardiness, social disengagement, and sleep disruptions,

underscore the significance of utilizing technology, especially machine learning and data analytics, for comprehensive analysis of big data. These tools reveal relationships that traditional methods may overlook, offering a more complete perspective.

2. Literature Survey:

G. Jackson-Koku, [1] has built machine learning models on Reddit posts using pre-trained models, censored material, and deep learning approaches in order to respond to the BDI using a variety of models. The BDI_Multi_Model outperforms state-of-the-art models by employing the optimal model for every query Overall depression prediction scores increased from 41.25% to 48.75%, with 84% of categories receiving favorable ratings. Gao, S.; Calhoun, V. Det.al [2] has described Future depression risk is predicted by the research using a cost function and the Decision Tree Algorithm with recursive binary splitting. The "Mental Checker" app evaluates a patient's risk of developing depression in the future by accounting for elements like stress, sickness, financial hardship, family and personal problems, and troubles in education. The program's usefulness was evaluated, and these factors aid in the classification of depression risks. Z. Jamil, D. Inkpen et.al [3] has proposed Using machine learning and natural language processing algorithms on social media data, the research proposes an effective way to identify depression and forecast its course. The proposed technique integrates behavioral analysis, topic modeling, and sentiment analysis to greatly improve the perfection of depression categorization when compared to previous methods. The study emphasizes the potential uses of social media data to diagnose mental health problems early on and have significant advantages for public health. [4]. SEM was used in the study to investigate the connections between student sadness, internet addiction, and Community support. The outcome says that Community support completely mediates the link between depression and internet addiction, emphasizing the necessity of addressing both variables to lessen depression in college students. [5]. Researchers applied NLP to Facebook data to develop a Thai community algorithm for detecting sadness, using quantitative engagement metrics like post frequency, interactions, privacy settings, and posting timing. Among three machine learning models—Random Forest, Deep Learning, and SVM—Random Forest proved most accurate, achieving an 84.6% precision in predicting depression among participants. [6].With great accuracy and efficiency when processing huge datasets, the study used SVM for depression anticipation using machine learning. It was discovered that word embeddings and TF-IDF are efficient showcase techniques for selecting that enhance prediction results. The study demonstrates how machine learning and digital footprints could be employed to track mental health, combining psychological and linguistic factors to increase forecast accuracy. [7].The study uses social media postings and BDI answers to predict depression using a multimodal deep learning strategy that integrates transformers, LSTMs, and hierarchical attention networks (HAN). In BDI_Multi_Model, the use of three LSTMs, two transformers, and five HANs greatly increased accuracy measures such as Average Hit Rate (AHR) and Depression Category Hit Rate (DCHR). This method appears to be promising for tracking various depression patterns in the broader community.[8].Using BERT/RoBERTa for similarity-based techniques and Top2Vec for topic-based tactics, the study used deep learning to categorize user posts by BDI categories. Accuracy in identifying certain depressive symptoms such as changes in appetite and suicidal thoughts was improved by using Bi-LSTM and HAN models customized to each BDI question. All things considered, the hybrid approach outperformed more contemporary algorithms in precisely recognizing and classifying depressive

symptoms from user postings. [9].The study

focused on discrete classifications and used SVM to evaluate depression severity based on clinical and neuroimaging data. It found neuroimaging indicators associated with depression and demonstrated machine learning's capacity for clinical diagnosis, indicating implications in therapeutic mental health treatments.[10].Using Random Forest, Deep Learning, and SVM, the study examined sadness through Facebook postings; Random Forest achieved an accuracy of 84.6% and an F-measure of 88.9%. Sentiment polarity and the frequency of posts on particular days were significant determinants; nevertheless, there were limitations, such as a limited sample size and difficulties with language translation.

3. Methodology:

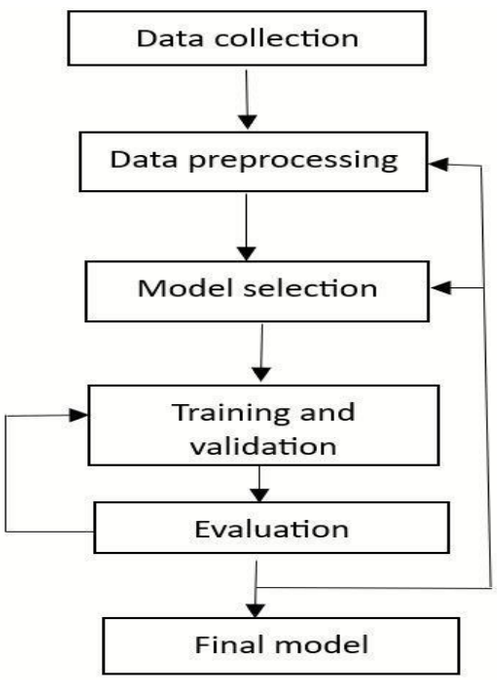

Fig 3.1: Flow diagram

The prediction of depression in students can be achieved using machine learning techniques such as Random Forest, Decision Tree, and Logistic Regression. By utilizing datasets with various parameters, these methods enable accurate prediction and analysis of results. The data for this analysis is collected in real time through Google Forms filled out by students.

Data-collection: Research in the natural sciences, business, health, and social sciences depends on data collection, which merges information from different sources to fully understand subjects or occurrences. Real-time data collection is often utilized, with tools like Google Forms playing a key role in gathering this information. This process aids in answering research questions, testing theories, and assessing results using both qualitative and quantitative approaches.

Data-preprocessing: Anxiety and stress are the integral parts of the modern world and affect people all over the globe. According to the current research, technology can also obtain finer health information when compared to conventional approaches. Random forest and decision tree grouping of forms of ML come into play to preprocess health care data by cleaning, transforming, and tabulating the raw inputs into useful data forms. To this end, this approach reveals patterns in data that conventional methods of data analysis cannot observe. These are the ML techniques that play vital role in the assessment of the disease probabilities and generation of recommendations in the context of mental health.

Model selection: entails selecting the machine learning method that is best suitable or framework Based on the dataset and problem at hand, aiming to achieve optimal performance and generalization. It involves evaluating multiple models to select the one best suited for the task.

Training: Prepared data is supplied to machine learning models during data training in order to maximize performance and modify model parameters for precise predictions or classifications. It is an important step in creating reliable and efficient AI systems.

Evaluation: use metrics such as recall, accuracy, and precision, and F1-score to compare a matching the forecast of a machine learning model with real outcomes with respect to evaluate the model's performance. It establishes how effectively the model adapts to fresh, untested data.

Final model: is the machine learning method chosen so that demonstrates the best performance after training and evaluation, ready for deployment in real-world applications. It represents the culmination of model development and optimization efforts.

3.1Algorithm:

3.1.1 Random-forest: Using random selections of data and characteristics, The Random Forest ensemble learning approach is Used to develop several decision trees which then combine their predictions. It increases accuracy and decreases over fitting in tasks involving regression and classification.

3.1.2: Decision tree: A model called a decision tree separates data into groups on basis of the values of its features. At the leaf nodes of the tree, these branches lead to judgments or predictions. It's straightforward to comprehend and useful for Regression and classification scenarios, but it has the capacity to over fit.

3.1.3 Logistic regression: By fitting data to a logistic curve, the statistical model known as logistic regression is needed for binary classification to estimate the likelihood of a given result. Using a logistic function, it calculates the connection between a dependent binary variable and one or more independent variables.

4.Results and discussion:

Different performance qualities are revealed when depression is predicted utilizing Random Forest together with Logistic Regression, and Decision Tree models. The ensemble method used by Random Forest minimizes overfitting and guarantees high accuracy and resilience by averaging many trees. Interpretability is offered by logistic regression, but it functions best with information that can be divided linearly and may have trouble with patterns that are complicated. Although interpretable, decision trees can overfit. In general, Random Forest is a top option for depression prediction across a variety of datasets due to its superior accuracy and generality.

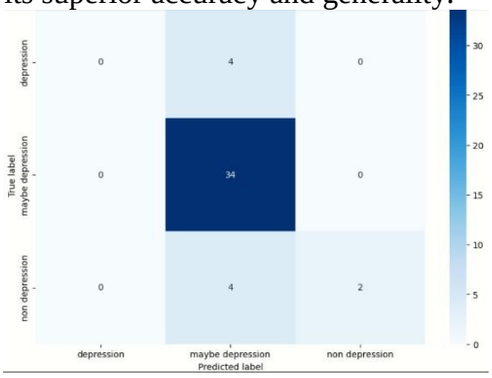

Fig 4.1 Confusion matrix

The image depicts a confusion matrix, which is commonly used to evaluate the performance of a classification model. The model accurately predicted "maybe depression" 34 times, indicating a strong performance for this category. For "non depression," the model made accurate predictions only 2 times.

5.Conclusion:

Given their capacity to manage intricate interactions and provide lucid probabilistic insights, machine learning methods such as Random Forest, Decision Trees, and Logistic Regression exhibit promise in the prediction of depression. Despite Decision trees functioning well for categorical data analysis, Random Forest is best at capturing non-linear correlations, and Logistic Regression is best for clear model interpretation. By combining these methods, prediction accuracy is improved, which aids in identifying those who are at-risk and provides opportunities for

feature augmentation and parameter refining in clinical applications.

References:

1 G. Jackson-Koku, “Beck Depression Inventory,” OCCMED, vol. 66, no. 2, pp. 174–175, Feb. 2016, doi: 10.1093/occmed/kqv087.

2 Gao, S.; Calhoun, V.D.; Sui, J. “Machine learning in major depression”: From classification to treatment outcome prediction. *CNS Neurosci. Ther.* **2018**, *24*, 1037–1052. [CrossRef] [PubMed].

3 Z. Jamil, D. Inkpen, P. Buddhitha, and K. White, ‘‘Monitoring tweets for depression to detect at-risk users,’’ Assoc. Comput. Linguistics, Vancouver, BC, Canada, Aug. 2017, pp. 32–40. [Online]. Available: https://aclanthology.org/W17-3104/

4 M. Gaikar, J. Chavan, K. Indore, and R. Shedge, “Depression Detection and Prevention System by Analysing Tweets,” *SSRN Electron. J.*, pp. 1–6, 2019, doi: 10.2139/ssrn.3358809.

5 A. D. Vergaray, J. C. H. Miranda, J. B. Cornelio, A. R. L. Carranza, and C. F. P. Sánchez, “Predicting the depression in university students using stacking ensemble techniques over oversampling method,” *Inform Med Unlocked*, p. 101295, Jul. 2023, doi: 10.1016/j.imu.2023.101295.

6 M. S. Zulfiker, N. Kabir, A. A. Biswas, T. Nazneen, and M. S. Uddin, “An in-depth analysis of machine learning approaches to predict depression,” *Current Research in Behavioral Sciences*, vol. 2, Jul. 2021, doi: 10.1016/j.crbeha.2021.100044.

7 Cummins N, Scherer S, Krajewski J, Schnieder S, J Epps, and T. F. Quatieri, “A review of depression and suicidality risk assessment using speech analysis,” Speech Communication, vol. 71, pp. 10–9, 2015 .

8 N. Cummins, J. Epps, M. Breakspear, and R. Goecke, “An investigation of depressed speech detection: Features and normalization,” in Twelfth Annual Conference of the International Speech Communication Association, 2011.

9 Tao, X.; Chi, O.; Delaney, P.J.; Li, L.; Huang, J. Detecting depression using an ensemble classifier based on Quality of Life scales. *Brain Inform.* **2021**, *8*, 2. [CrossRef].