STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 9, Issue 3, 2022
DOI Journal: 10.37591/JoAIRA

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Review JoAIRA

Analysis of Machine Learning Algorithms for the Detection of Mental Health Problems in Employees

Trailokya Raj Ojha*

Abstract

Mental health-related issues are more critical issues in the workplace. Mental health has been a major and challenging issue, especially for working professionals. A bad work environment can cause a variety of physical and physiological problems as well as decreased productivity. People are reluctant to ask for help from mental health specialists, according to studies. The fundamental reason for this is the stigma attached to mental illness. Working professionals therefore have a higher risk of mental health issues. To analyze the consequences of mental health in professionals, in this study, we have analyzed different factors related to personal, professional, and family history using different machine learning algorithms. The algorithms used for the analysis of mental health problems are Naïve Bayes, Logistic Regression, Adaptive Boosting (AdaBoost), Random Forest, and Decision Table. The result obtained from the model training shows that the best performance was achieved using the Random Forest algorithm with an accuracy score of 83.3%. In terms of precision, the AdaBoost algorithm performed best with a precision score of 84.6%. Mental health issues at the workplace are a critical issue and it is believed that the findings from this study will help to aid in medical health to minimize mental health issues at the workplace.

Keywords: Depression, mental health, stress, employees, tech industry, machine learning

INTRODUCTION

The most prevalent mental condition is depression, which affects many people either directly or indirectly due to a friend or family member. There is usually misunderstanding about depression, including what it is and how it varies from other emotions. A person's level of mental health acts as a gauge for how to treat their illnesses effectively. It is crucial to monitor the mental health professional files of various groups to anticipate any health-related anomalies. According to Hans Selye, stress can be defined as "a nonspecific response of an organism in exposure to a demand or a change in the physical situation" [1]. A World Health Organization (WHO) statistic states that psychological or neurological hazards will affect a person at some point in their lifetime [2]. It is essential to assess the mental health of various categories at various points in time to prevent major sickness.

A pattern called occupational stress arises as a result of mental, emotional, behavioral, and physical reactions to unfavorable aspects of the structure, surroundings, and content of the work [3]. Occupational stress, a more recent issue, has drawn attention as a serious health concern [4]. Workplace stress is a major concern for the smooth operation of any organization.

Stressful situations arise as a result of the employees' need to work under time constraints to

***Author for Correspondence**
Trailokya Raj Ojha
E-mail: trailokyaro@nec.edu.np

Assistant Professor, Department of Computer Science and Engineering, Nepal Engineering College, 44800 Bhaktapur, Nepal

Received Date: September 15, 2022
Accepted Date: December 28, 2022
Published Date: January 05, 2023

Citation: Trailokya Raj Ojha. Analysis of Machine Learning Algorithms for the Detection of Mental Health Problems in Employees. Journal of Artificial Intelligence Research & Advances. 2022; 9(3): 12–24p.

finish the task and deliver on schedule. The key reasons for the high levels of stress experienced by IT professionals are role conflict, corporate culture, role overload, intercultural staff, customer demand, long workdays, and a lack of managerial support. In this study, we have examined various machine learning techniques in various contexts to make it simple to comprehend the aspects influencing the mental health of workers in tech companies. Based on the data provided by the employees in the questionnaire, various machine learning algorithms such as Nave Bayes, Logistic Regression, Adaptive Boosting (AdaBoost), Random Forest, and Decision Table were applied to estimate the likelihood of the appearance of mental health problems.

The data set adopted in this study was taken from the Kaggle repository [5]. Usually, data taken from a secondary source are not ready to use by machine learning algorithms. To make data usable by machine learning algorithms, different data pre-processing techniques have been used. After making the data set ready for processing, it was applied to machine learning models, and the result was compared to predict the best model for the detection of mental health problems in employees.

RELATED WORKS

According to world health organization, a mental health is defined as "a condition of well-being in which a person is aware of his or her skills, can manage everyday pressures, can work productively, and can contribute to his or her community" [6]. A machine learning approach was employed in a study by Moon *et al.* to predict Bangladesh's employment downturn [7]. Nave Bayes, K Neighbors, Random Forest Classifier, and Random Forest Regression were employed by the authors.

The study evaluated the effectiveness of eight different machine learning algorithms that categorize information into different mental health conditions. According to their findings, the Multiclass Classifier, Multilayer Perceptron, and LAD Tree are the three classifiers that produce results that are more precise than the others [8]. According to Passos *et al.*, it is possible to determine whether a person will attempt self-destruction by using segment factors and clinical variables linked to self-destruction [9]. They employed three different MATLAB-based algorithms: LASSO, SVM, and RVM. The precision range was frequently between 65 and 72% and RVM produced the result with the highest accuracy (72%).

It was also found that using ensemble classifiers significantly improved mental health prediction performance, with a 90% accuracy rate [10]. The study has focused on how machine learning can help with the identification and diagnosis of mental diseases like schizophrenia, depression, and Alzheimer's disease. A rainfall prediction model might be created with the help of AI and the LSTM approach, according to Salehin *et al*. [11]. Deep learning methods are used because they produce more accurate results. Six variables were mentioned in their article. The accuracy of the evaluation of all the data is 76%. Salehin *et al*. predicted the severity of depression caused by heavy cell phone use [12]. To identify depression, they have employed machine learning techniques such as linear regression and decision trees.

Hassan *et al*. researched sentiment analysis and its application to depression analysis [13]. They have used the Support Vector Machine (SVM), Nave Bayes, and Maximum Entropy classifier approaches to distinguish between the results of sentence-level sentiment analysis. The selection of features was done using a voting approach. The results showed that SVM was a significantly more accurate technique than Naive Bayes and Maximum Entropy classifiers, with an accuracy of 91%. In a study conducted by Mutalib *et al*., they used the methods of neural networks, decision trees, naive bayes, support vector machines, and logistic regression to predict mental health issues in college students [14]. The most accurate models for stress, depression, and anxiety are the Support Vector Machine, Decision Tree, and Neural Network, in that order. Shetu *et al*. predicted student performance using educational data mining [15]. Based on the individual's academic position and the academic environment in which they are situated, a comprehensive prognosis was made.

To analyze occupational stress, Yadav and Hashmi conducted a study using a support vector machine, a neural network, a decision tree, and a random forest algorithm [16]. For training and testing, a stratified 10-fold cross-validation was employed. 60% accuracy rates, 80% sensitivity rates, and 60% specificity rates were attained using the suggested model. To examine the Effective Factors of job stress, Khaleghi *et al*. performed the correlation coefficient test and progressive multivariate regression [17]. Ghani *et al*. addressed the link between occupational stress and health using demographic characteristics such as age, gender, highest academic qualification, marital status, employment tenure, and job title [18]. Aytac used multiple regression analysis to evaluate the level of stress caused by psychological risk factors [19].

METHODOLOGY

In this study, a prediction model is developed to identify mental health disorders in workers in the workplace using a machine learning approach. The primary processes in the designed model are data collection, pre-processing, and data manipulation utilizing data mining techniques. We have used various machine learning classification techniques like the Naïve Bayes, Logistic regression, Adaptive Boosting (AdaBoost), Random Forest, and Decision Table. The proposed working model is depicted in Figure 1.

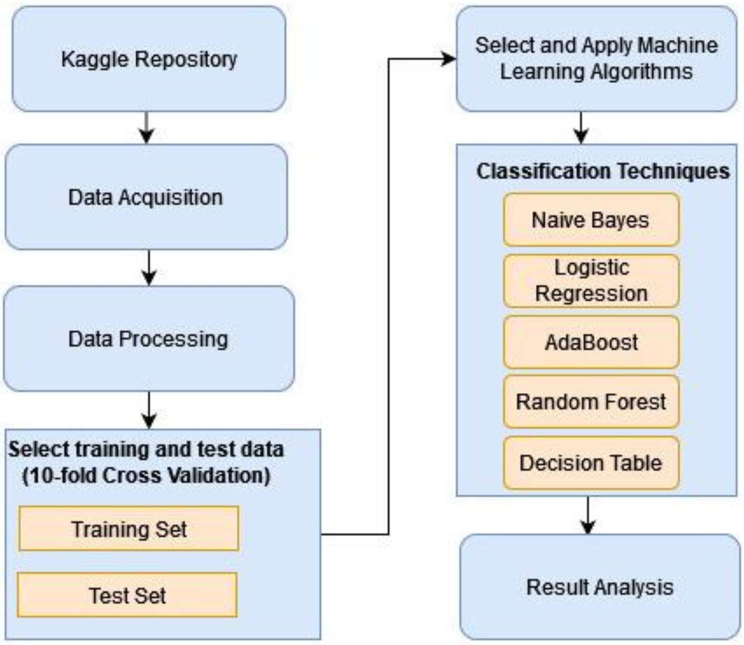

Figure 1. Flowchart of the working methodology.

Dataset Description

The dataset was taken from the Kaggle repository to conduct this study [5]. It contains 1259 values with 26 attributes and one class variable. The class variable (treatment) contains two values, 'Yes' and 'No'. 'Yes' defines whether the employee has ever gone for mental health treatment and 'No' defines whether the employee has not gone for mental health treatment. A detailed description of dataset attributes with possible values is presented below:

- *Timestamp* (Predictor Variable): the time when the participant completed the survey.
- *Gender* (Predictor Variable (Male, Female, Other)): describes the gender of the participant.
- *Age* (Predictor Variable (21 to 87)): age of the participant.
- *Country* (Predictor Variable (country name)): describes the participant's country of work.
- *State* (Predictor Variable (state name)): describes the state in which the participants are working.
- *Self_employed* (Predictor Variable (Yes, No, NA)): describes whether the participant is self-employed or not.

- *Family_history* (Predictor Variable (Yes, No)): does the participant have a family history of mental illness?
- *Work_interfere* (Predictor Variable (often, rarely, never, sometimes, NA)): does the participant have a family history of mental illness?
- *No_employees* (Predictor Variable (1–5, 6–25, 26–100, 100–500, 500–1000)): number of employees working in the company.
- *Remote_work* (Predictor Variable (Yes, No)): whether the participant works remotely at least 50% of the time.
- *Tech_company* (Predictor Variable (Yes, No)): the participant is working in a tech company or not.
- *Benefits* (Predictor Variable (Yes, No, Don't know)): describes whether the employer provides mental health benefits or not.
- *Care_option* (Predictor Variable (Yes, No, Not sure)): does the employer provide mental health care or not?
- *Wellness_program* (Predictor Variable (Yes, No, Don't know)): whether the employer discussed mental health as a part of the employee's wellness program.
- *Seek_help* (Predictor Variable (Yes, No, Don't know)): does the employer have a plan for how to seek help in mental health?
- *Anonymity* (Predictor Variable (Yes, No, Don't know)): whether privacy is maintained when you take mental health treatment or not.
- *Leave* (Predictor Variable (very easy, very difficult, somewhat easy, somewhat difficult, don't know)): describes how easy it is to take leave for treatment.
- *Mental_health_consequence* (Predictor Variable (Yes, No, Maybe): whether it affects negatively if participants discuss mental health issues with the employer.
- *Phys_health_consequence* Predictor Variable (Yes, No, Maybe): whether it affects negatively if participants discuss physical health issues with the employer.
- *Coworkers* (Predictor Variable (Yes, No, some of them)): whether the participant wants to discuss mental health issues with colleagues?
- *Supervisor* (Predictor Variable (Yes, No, some of them)): Does the participant want to discuss a mental health issue with supervisors?
- *Mental_health_interview* (Predictor Variable (Yes, No, Maybe)): Describes whether the participant wants to talk about their mental health issue in an interview.
- *Phys_health_interview* (Predictor Variable (Yes, No, Maybe)): describe whether the participant wants to talk about their physical health issue in the interview.
- *Mental_vs_physical* (Predictor Variable (Yes, No, Maybe)): do employers think mental health is as serious as physical health?
- *obs_consequence* (Predictor Variable (Yes, No)): does the participant know about the negative impact of mental health in the workplace?
- *Comments* (Predictor Variable): open comments from participants.
- *Treatment* (Response Variable (YES, NO)): response variable that describes the status of mental health treatment.

The data distribution according to gender and mental health treatment status is shown in a pie chart in Figure 2(a) and (b).

Data Pre-processing

The final result is frequently adversely affected by the noisy or unbalanced data set. To balance the data set, the data preparation technique should be used. By preparing data, we enable machine learning models to use raw data. A class variable and 26 attributes make up the data set used in this study. At this level, we first determined which attributes are relevant to predicting the outcome and which are not. After carefully examining each attribute, it was decided to remove the timestamp, country, state, and comment attributes because they do not significantly contribute to the goal. The numbers in the age column are corrected and we have adjusted null values whenever they appear in the dataset. Some entries in the age column had negative values, below 18, or greater than 100. As a result, these values are removed. Since WEKA cannot process distinct sorts of values, all the values were ultimately transformed into nominal values. The data set is well balanced, with 637 entries in the treatment column that are marked "Yes" and 622 that are marked "No".

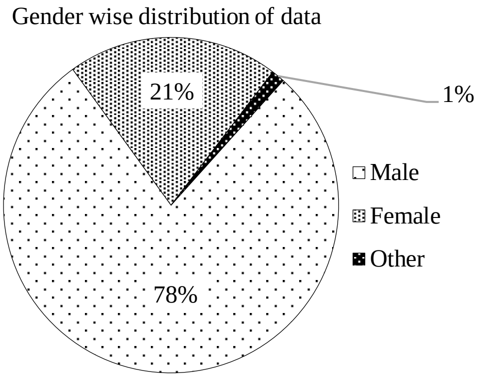

Figure 2. (a) Gender-wise data distribution.

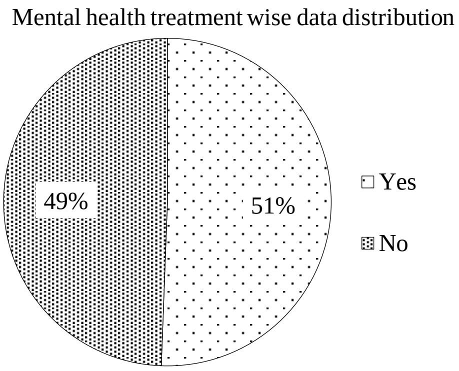

Figure 2. (b) Treatment-wise data distribution.

The next stage is to build a model after dealing with the unbalanced dataset and finishing data pre-processing. Training and testing data are separated from the preprocessed data set. In this work, we employed a 10-fold cross-validation method to improve the outcome. 90% of the entire training dataset is randomly chosen for use as training data in the 10-fold cross-validation process, and the remaining 10% is used as test data. After dividing the dataset, we used classification, clustering, and association methods to train the model. The following is a description of the various machine learning techniques that were utilized to train the model.

Algorithm Description

Naïve Bayes

Learning is greatly facilitated by the Naive Bayes classifier's assumption that attributes are independent of class. Naive Bayes frequently outperforms more sophisticated classifiers in practice, even though independence is frequently a false assumption. The Bayes theorem offers the conditional probability of an event occurring in comparison to an already occurring event. A naïve Bayes classifier cannot perform if the data set contains many attributes. But it performs better at handling small datasets that need less training data, is highly scalable, adept at handling both continuous and discrete data, and is insensitive to irrelevant features [20]. The precision of the Nave Bayes classifier is not directly related to the degree of dependencies obtained as class-conditional information between different characteristics [21].

Logistic Regression

When the dependent variable is dichotomous, logistic regression is the appropriate regression strategy to utilize. It is a predictive analysis technique. When the dependent variable (target) is categorical, it is used. It approximates using the maximum likelihood estimation technique. Throughout the instruction session, n unique multiple decision trees are generated. The final decision tree is formed, and the data record that was first incorrectly recognized is given top priority. Based on this, the output from these is gathered and sent to the subsequent decision model. This method iterates and repeats itself till we achieve the required base of learners that was proposed to be established in the beginning [22].

Adaptive Boosting (AdaBoost)

The AdaBoost algorithm is one of the most typical and widely used ensemble learning methods. Boosting is a technique that combines all ineffective classifiers into one powerful classifier. Multiple weak learning entities are used by AdaBoost and combined into a single model. Stumps or single split decision trees are used in this strategy. The AdaBoost algorithm was developed by Freund and Schapire, who claimed it was more practical and simpler than the prior boosting techniques [23]. In the initial iteration, each stump had the same weight. To improve the previous learner, incorrect classifications are given more weight than the right ones. Until we reach the necessary base of learners that were suggested to be created at the beginning, this process iterates and repeats itself [22].

Random Forest

Using randomly chosen qualities at each node, the random forest classification algorithm constructs a tree. To build a training data set using a random way drawing with substitute N values, where N is equal to the size of the initial training set, bagging was employed for each feature/feature combination that was chosen [24]. A flexible machine learning algorithm is Random Forest. Both classification and regression issues can be handled by it. It performs a respectable job and can also assist with dimension reduction, missing values, extreme values, and other crucial data exploration procedures. The random forest classification technique uses the Gini index as an optimization criterion to quantify the degradation of a class attribute. By picking a case (pixel) at random and indicating that it belongs to a certain class, the Gini index for a particular training set S can be expressed as follows:

$$\sum\sum_{j\neq i}(f(X_i,S)/|S|)(f(X_j,S)/|S|)$$

Where, $f(X_i, S)/|S|$ is the probability showing the selected case belongs to class X_i.

Whenever a tree using a feature combination and certain training data expands to its maximum depth. Trees that are mature enough are not pruned [25].

Decision Table

Decision tables (DTs) present an alternative method for describing rule-based classification models in a user-friendly way [26]. DTs are represented in tabular form and are used to describe and analyze decision circumstances (such as the assessment of credit risk), in which the existence of several criteria affects the execution of a series of actions [27]. In our context, conditions correspond to the outcome of mental health treatment or not. A decision table consists of four quadrants as shown in Table 1. The horizontal line is used to divide the table into a condition part (above) and an action part (below). The vertical line is used to separate subjects (left) from entries (right).

Table 1. Quadrants of the decision table.

Condition subjects	Condition entries
Action subjects	Action entries

The criteria that are important to the decision-making process are the condition subjects. They stand for the characteristics of the rule preceding for which information is required to categorize whether or not the participant has a stroke. For each condition subject (attribute), each condition entry defines a relevant subset of values (referred to as a state), or it contains the dashed sign ('-') if the value is not important for that column. The values allocated to the associated action subject (class) are then stored in the action entries, with the "x" entry designating the value that corresponds to a certain set of circumstances. Thus, each column in the DT's entry section has a classification rule.

RESULTS

In this section, the result of the study is summarized. The main objective of this study was to understand the significance of depression analysis in working employees and to perform a comparative analysis of algorithms in search of the algorithm that would work best. To obtain the result, WEKA 3.8.6 environment was used [28]. A ten-fold cross-validation technique was used to obtain the training and test set of data and different classification algorithms were used to train the model. The performance of the different classifiers is measured in terms of different accuracy metrics such as accuracy, precision, recall, f-measure, and receiver operating characteristic (ROC) curve. These performance metrics for the different classification algorithms are illustrated in Table 2.

Table 2. Detailed performance of ML algorithms.

Algorithm	Accuracy	Precision	Recall	F-Measure
Naïve Bayes	80.0%	80.0%	80.0%	80.0%
Logistic Regression	82.7%	83.1%	82.7%	82.6%
AdaBoost	82.2%	84.6%	82.2%	81.8%
Random Forest	83.3%	83.8%	83.3%	83.2%
Decision Table	82.7%	83.6%	82.7%	82.5%

The result obtained from a simulation in WEKA shows that all the classification algorithms implemented in the study perform well with an accuracy above 80%. Among all, the Random Forest classifier performs the best with an accuracy of 83.3%. The other accuracy metrics such as precision, recall, and f-measure for the same classifier are recorded with the value of 83.8, 83.3, and 83.2% respectively. Similarly, the accuracy, precision, recall, and f-measure for the Naïve Bayes classifier are achieved with the value of 80.0%. The logistic regression classifier also has good performance with 82.7% of accuracy, 83.1% precision, 82.7% of recall value, and 82.6% f-measure score. The accuracy score of AdaBoost and Decision Table classifier is 82.2 and 82.7% respectively. The precision, recall, and f-measure values for the AdaBoost classifier are noted with the value of 84.6, 82.2, and 81.8% respectively. The Decision table classifier has 83.6, 82.7, and 82.5% precision, recall, and f-measure respectively.

In terms of accuracy, the Random Forest classifier performs the best with an accuracy score of 83.3% and in terms of precision, the AdaBoost classifier has the best performance (84.6%). If we compare the accuracy score, the maximum difference is found to be 3.3% between Naïve Bayes and Random Forest classifiers. The Logistic Regression and Decision Table classifier performed in the same way to predict mental health issues in employees with an accuracy of 82.7%. The comparison of the different performance metrics for the different classifiers is depicted in Figure 3.

The different performance metrics of the machine learning algorithms used in the study are represented in a line chart graphically in Figure 4.

If we analyze the receiver operating characteristic (ROC) curve for different machine learning algorithms, it shows that all the algorithms have performed well. The ROC curve for all five machine learning algorithms used in this study is shown in Figures 5–9.

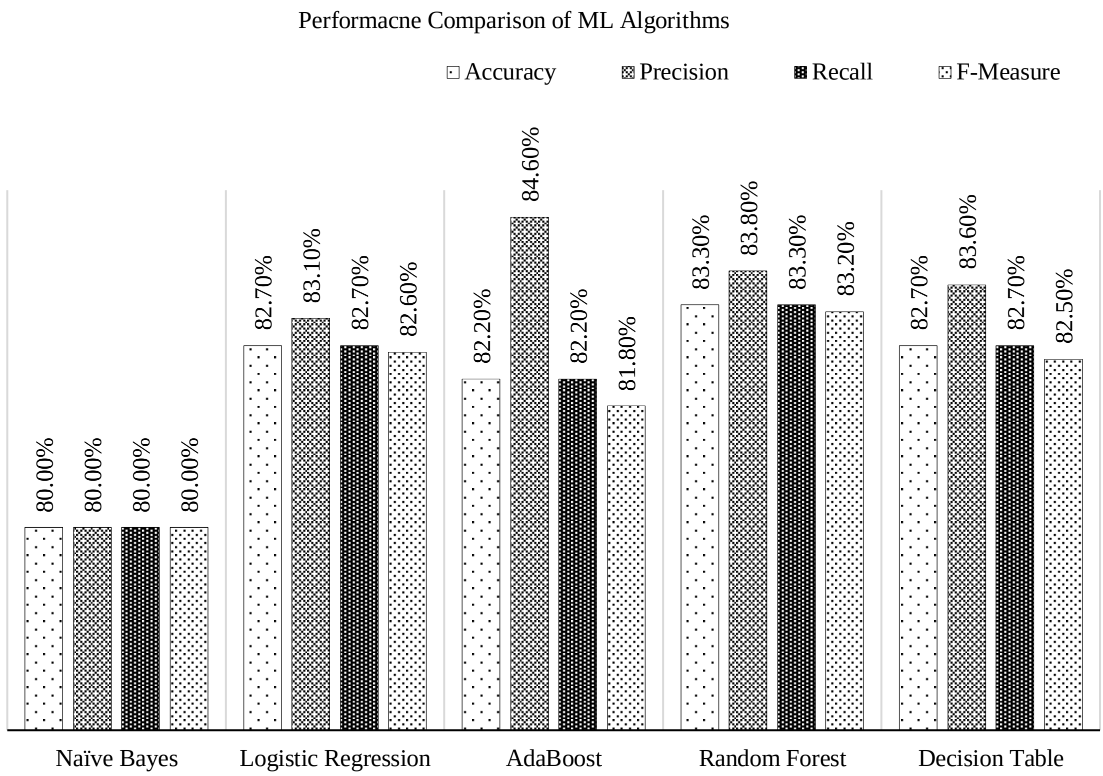

Figure 3. ML algorithms performance comparison.

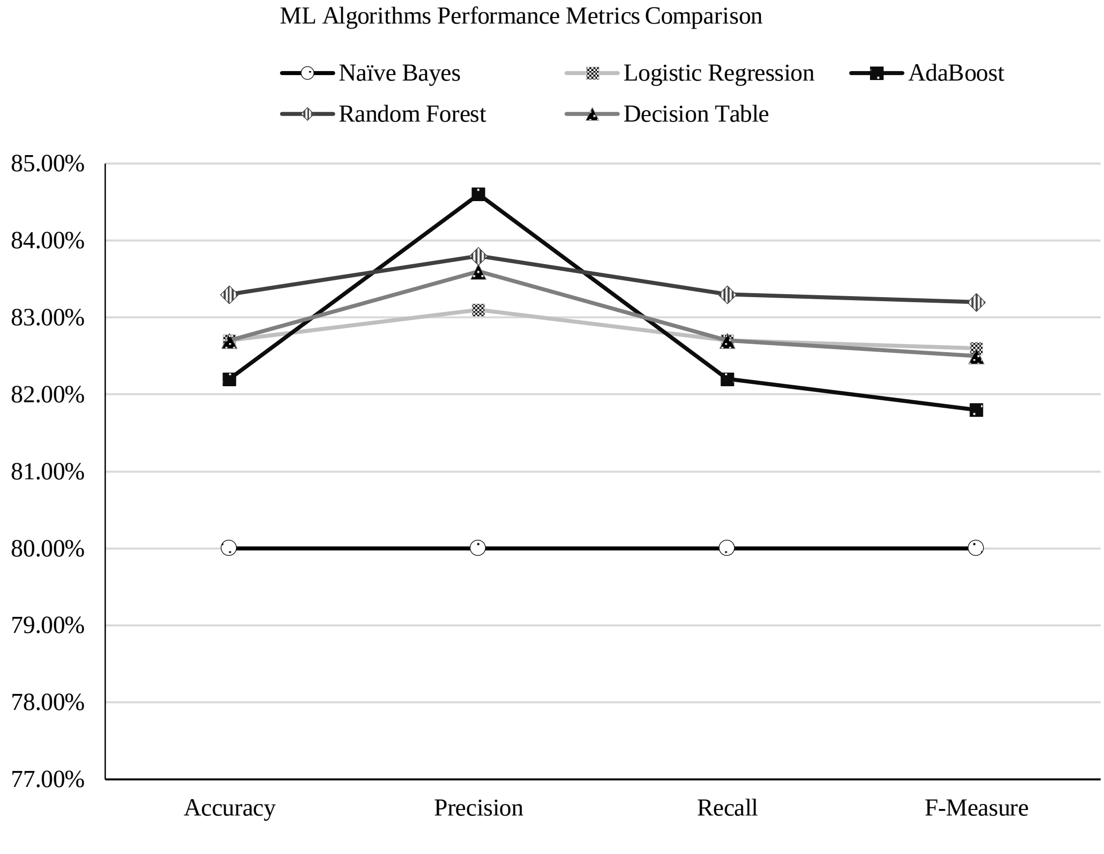

Figure 4. Performance metrics of different ML algorithms.

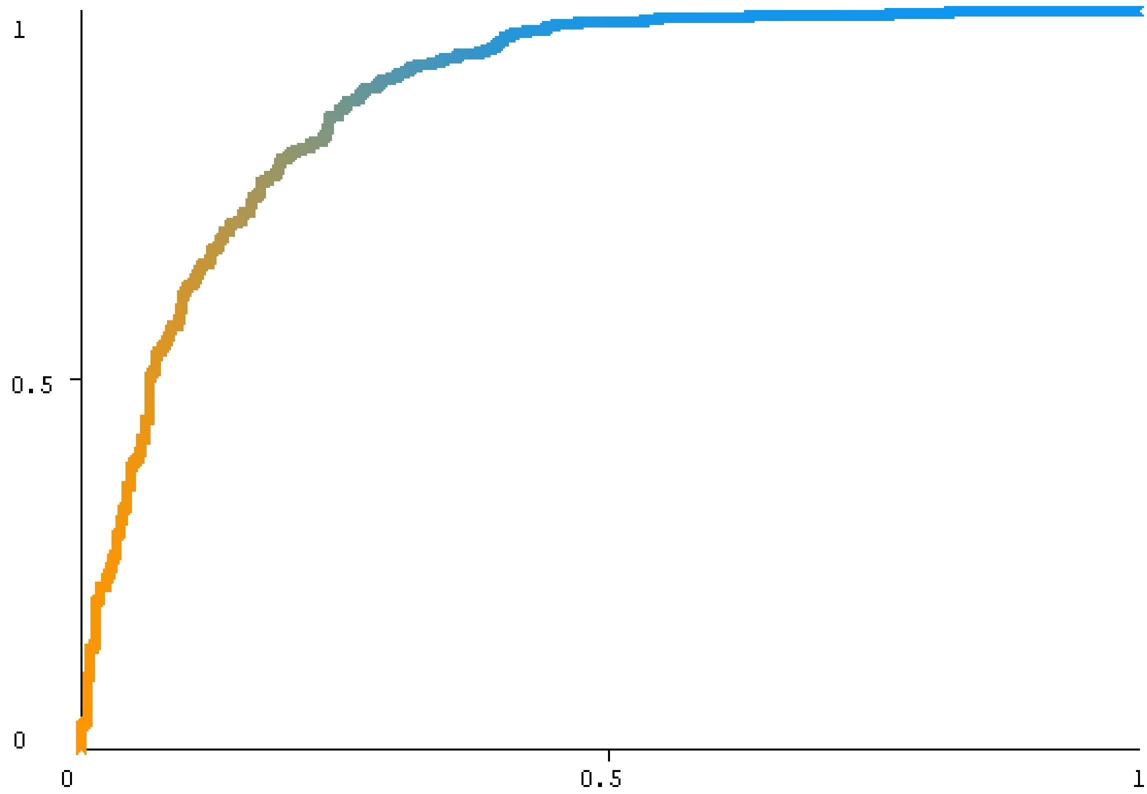

Figure 5. ROC curve for Naïve Bayes.

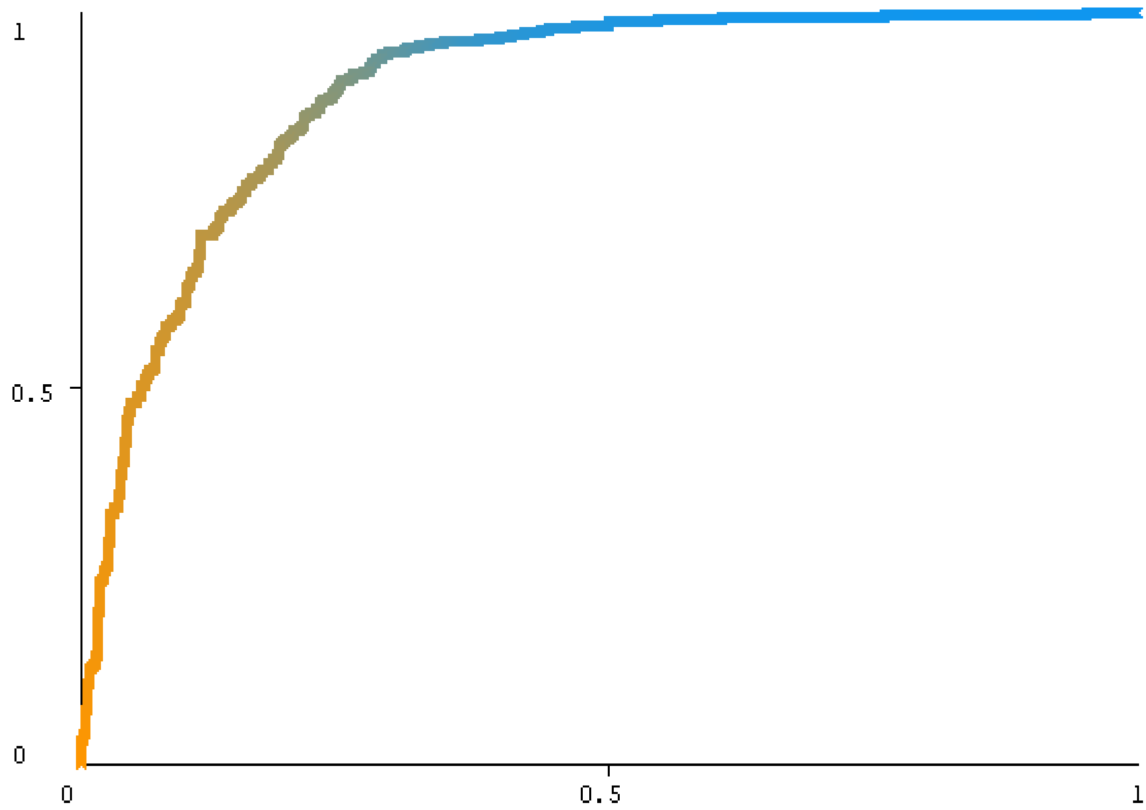

Figure 6. ROC for logistic regression.

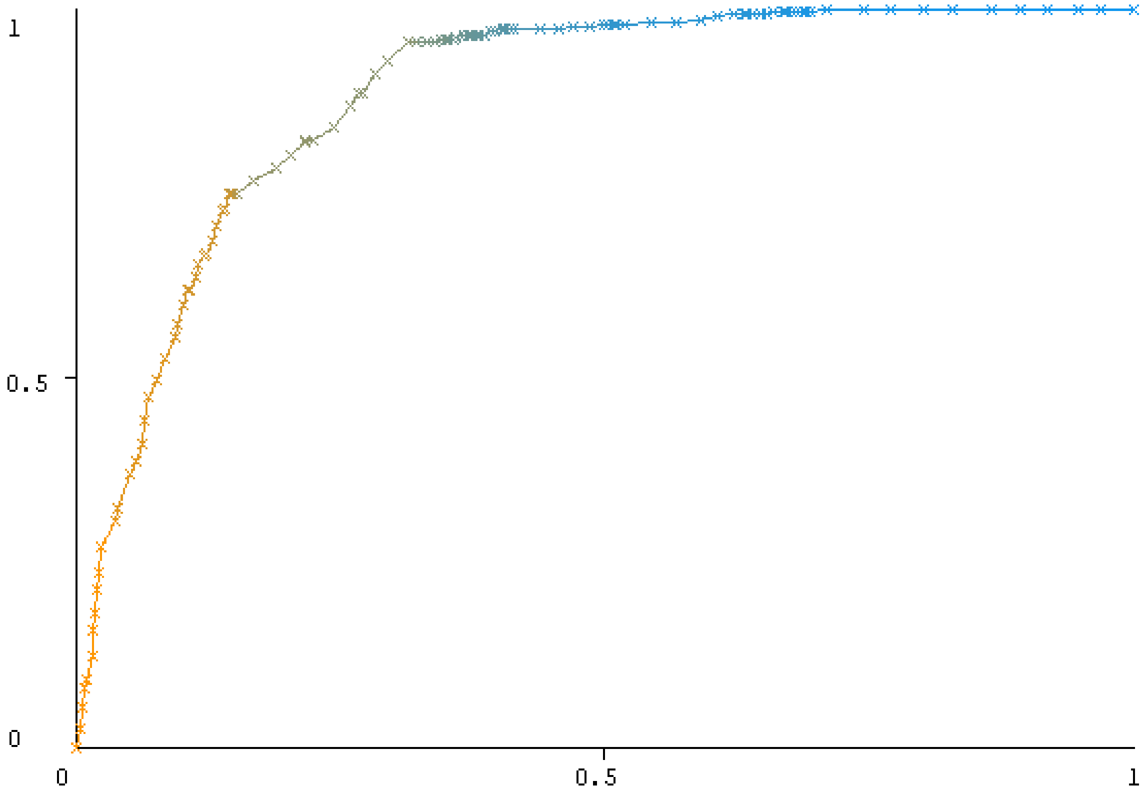

Figure 7. ROC curve for AdaBoost classifier.

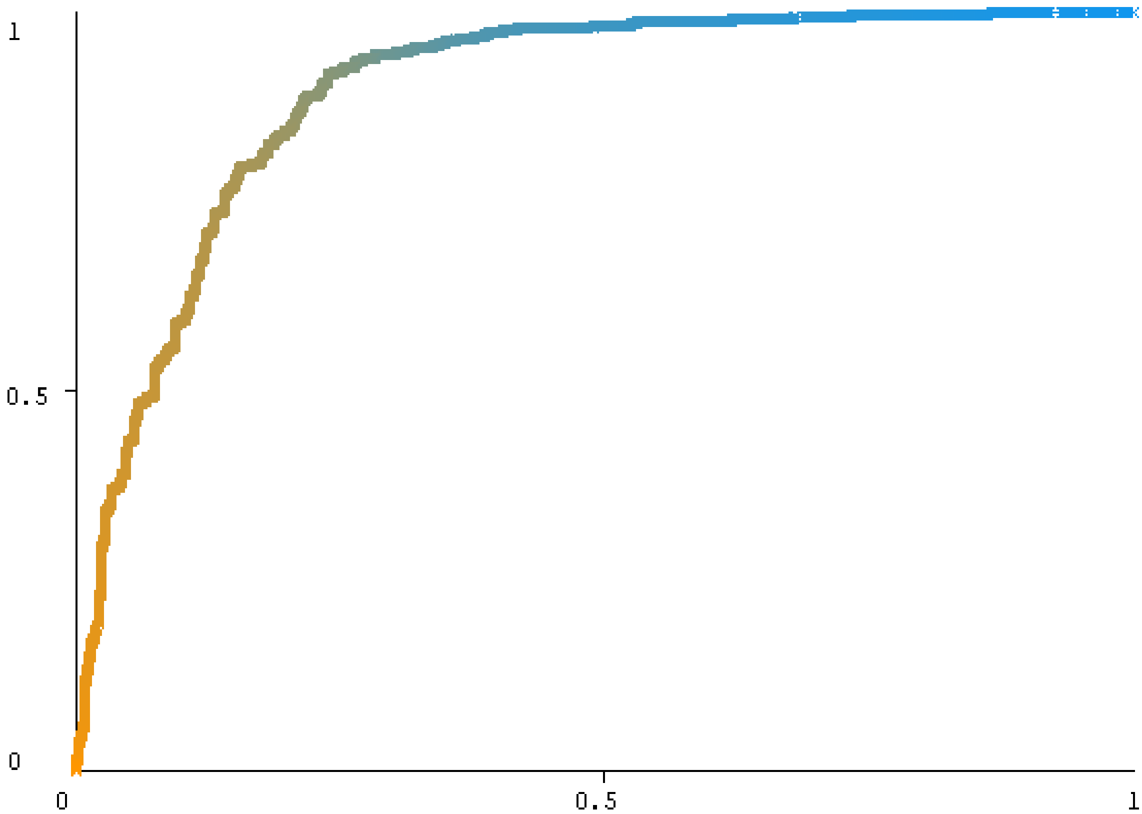

Figure 8. ROC for random forest.

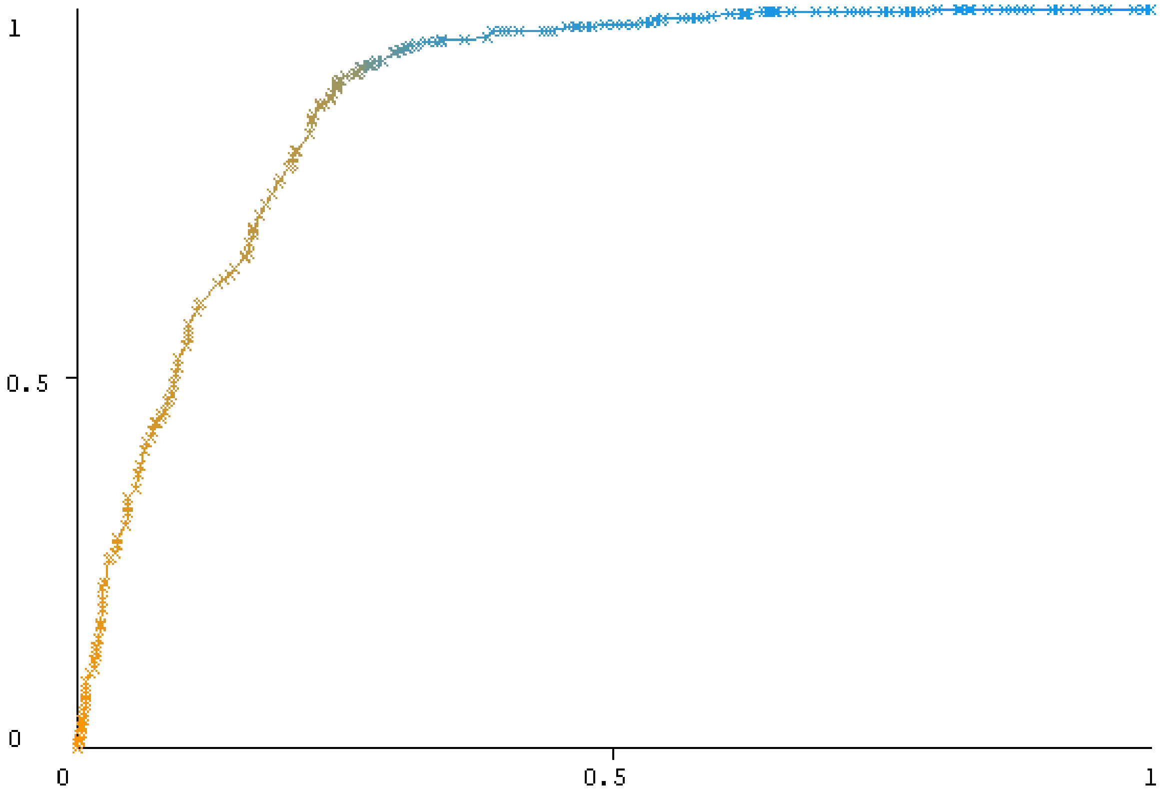

Figure 9. ROC for decision table.

CONCLUSION

Mental health is a crucial element for employees working in a workplace. Additionally, it influences how things will come out in terms of efficiency, productivity, financial and general terms. The first step in putting avoidance and clinical mediation strategies into practice is the early prediction and identification of depression. Similar to the machine learning model developed in this study, such models can offer a technically sophisticated approach to medical science for detecting and treating depression in working employees, which is mostly dependent on data that is reported by the patient themselves. In this study, we have used different machine learning algorithms such as Naïve Bayes, Logistic Regression, Adaptive Boosting (AdaBoost), Decision Table, and Random Forest to forecast the possibility of depression in employees. Different performance metrics are considered from the results. The result obtained from the model training shows that the Random Forest classifier performs better in terms of accuracy (83.3%) and AdaBoost performs better in terms of precision (84.6%) compared to the rest of the models. This model will significantly help the medical field predict possible depression in employees in any workplace, especially in the tech industry.

REFERENCES

1. Tan SY, Yip A. Hans Selye (1907–1982): Founder of the stress theory. Singapore medical journal. 2018 Apr; 59(4): 170.
2. Adamchik V. Data structures and algorithms in pen-based computing environments. 2011 IEEE Global Engineering Education Conference (EDUCON). 2011 Apr; 1211–1214.
3. European Commission. Guidance on Work-related Stress. Health and safety at work, Luxembourg, Directorate-General for Employment and Social Affairs, Unit D.6. 1999.
4. Allison GJ. Perceived and reported occupational stressors and coping strategies of selected community college business faculty members in Texas. Texas A&M University; 2004.
5. Mental Health at Workplace dataset. [Online]. https://www.kaggle.com/code/aditimulye/mental-health-at-workplace/data. [Accessed: September 01, 2022].

6. World Health Organization. Mental health: Strengthening our response. Geneva, Switzerland. World Health Organization; 2018. Available from: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-ourresponse.
7. Moon NN, Mariam A, Sharmin S, Islam MM, Nur FN, Debnath N. Machine learning approach to predict the depression in job sectors in Bangladesh. Current Research in Behavioral Sciences (CRBS). 2021; 2: 100058.
8. Sumathi MR, Poorna B. Prediction of mental health problems among children using machine learning techniques. Int J Adv Comput Sci Appl. 2016; 7(1): 552–557.
9. Passos IC, Mwangi B, Cao B, Hamilton JE, Wu MJ, Zhang XY, Soares JC, et al. Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. J Affect Disord. 2016; 193: 109–116.
10. Srividya M, Mohanavalli S, Bhalaji N. Behavioral modeling for mental health using machine learning algorithms. J Med Syst. 2018; 42(5): 1–12.
11. Salehin I, Dip ST, Talha IM, Rayhan I, Nammi KF. Impact on human mental behavior after pass through a long time home quarantine using machine learning. Int J Educ Manag Eng. 2021; 11(1): 41–50.
12. Salehin I, Talha IM, Moon NN, Saifuzzaman M, Nur FN, Akter M. Predicting the depression level of excessive use of mobile phone: decision tree and linear regression algorithm. In 2020 IEEE International Conference on Sustainable Engineering and Creative Computing (ICSECC). 2020 Dec; 113–118.
13. Hassan AU, Hussain J, Hussain M, Sadiq M, Lee S. Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. 2017 IEEE International Conference on information and communication technology convergence (ICTC). 2017 Oct; 138–140.
14. Mutalib S, Shafiee NSM, Abdul-Rahman S. Mental health prediction models using machine learning in higher education institution. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 2021; 12(5): 1782–1792.
15. Shetu SF, Saifuzzaman M, Sultana S, Yousuf R, Moon NN. Students performance prediction through education data mining depending on overall academic status and environment. In 3rd International Conference on Innovative Computing and Communication (ICICC-2020). 2020; 757–769.
16. Yadav SK, Hashmi A. An Investigation of Occupational stress Classification by using Machine Learning Techniques. Int J Comput Sci Eng. 2018; 6(6): 842–850.
17. Khaleghi A, Najafabadi MO, Lashgarara F. Effective Factors on Job Stress from Experts' Perception; a Case Study in Iranian Agriculture Engineering Organization. International Journal of Review in Life Sciences. 2015; 5: 94–99.
18. Ghani MZ, Ahmad AC, Ibrahim S. Stress among special education teachers in Malaysia. Procedia-Soc Behav Sci. 2014; 114: 4–13.
19. Aytac S. The sources of stress, the symptoms of stress and anger styles as a psychosocial risk at occupational health and safety: a case study on Turkish police officers. Procedia Manuf. 2015; 3: 6421–6428.
20. Shoily TI, Islam T, Jannat S, Tanna SA, Alif TM, Ema RR. Detection of stroke disease using machine learning algorithms. In 2019 IEEE 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). 2019 Jul; 1–6.
21. Rish I. An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence. 2001 Aug; 3(22): 41–46.
22. Bandi V, Bhattacharyya D, Midhunchakkravarthy D. Prediction of Brain Stroke Severity Using Machine Learning. Rev d'Intelligence Artif. 2020; 34(6): 753–761.
23. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci. 1997; 55(1): 119–139.
24. Breiman L. Bagging predictors. Mach Learn. 1996; 24(2): 123–140.

25. Pal M. Random forest classifier for remote sensing classification. Int J Remote Sens. 2005; 26(1): 217–222.
26. Wets G, Vanthienen J, Piramuthu S. Extending a tabular knowledge-based framework with feature selection. Expert Syst Appl. 1997; 13(2): 109–119.
27. Vanthienen J, Dries E. Illustration of a decision table tool for specifying and implementing knowledge based systems. Int J Artif Intell Tools. 1994; 3(02): 267–288.
28. WEKA. (2021). The Data Platform for AI. WEKA. [online] Available from: https://www.weka.io/.