



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Predicting Mental Health of Employees Using Machine Learning

Nanda C M, Nisarga M C, Sahana S Sulakhe, Tazimbanu Hubballi, Murgesh V Jambigi

U.G. Student, Department of Computer Engineering, Sri Taralabalu Jagadguru Institute of Technology, Ranebennur,
Karnataka, India

Associate Professor, Department of Computer Engineering, Sri Taralabalu Jagadguru Institute of Technology,
Ranebennur, Karnataka, India

ABSTRACT: Mental health concerns have become a prominent issue in 21st-century healthcare, largely due to widespread lack of awareness. This paper endeavors to raise awareness among the public about common mental health issues by leveraging classification machine learning algorithms. To achieve this, data was gathered from individuals of diverse demographics, including age, profession, gender, and lifestyle, through surveys containing questions commonly employed by psychologists to gain insight into their patients' challenges

KEYWORDS: Mental health prediction, employee mental health, machine learning for mental health.

I. INTRODUCTION

Modern lifestyles are causing different types of psychological health problems in many peoples[2]. Mental well-being is critical to an individual's health and quality of life. It encompasses emotional, psycho-logical, and social dimensions, making it a complex and multifaceted construct. Traditionally, assessing and forecasting mental well-being has relied on self-report measures, clinical evaluations, and surveys, which can be subjective, time-consuming, and limited scope[3]. Psychological blems like anxiety, depression and stress have some overlapping features, for example, a person feels low and lonley in all three. Generally, psychiatrists assess anxiety depression and stress through questionnaire such as DASS21 and DASS42 because people suffering with anxiety, depression and stress are often not open to sharing their feelings with doctors, relatives or friends[2].

The World Health Organization (WHO) has observed that depression is the most prevalent mental disorder affecting more than 300 million people worldwide, and the severity of the issue has led many health researchers to focus their studies in this area. Differentiating anxiety, depression and stress from one another is problematic for machines; hence, an appropriate learning algorithm is required for an accurate diagnosis. According to WHO, a healthy person possesses a healthy brain along with physical wellness. Many researchers have worked on predicting anxiety and depression with machine learning algorithms, such as Random Forest Tree (RFT), the Support Vector Machine (SVM) and the Convolution Neural Network (CNN) for the collection and subsequent classification of data from blog posts. For encoding the text, various techniques have been used, that is topic modelling, Bag-of-Words (BOW) and Term Frequency–Inverse Document Frequency (TF–IDF). Moreover, Python programming has been used for modelling experiments, with the best results among all the classifiers[2] being produced by the CNN, whose accuracy and recall scores were found to be 78% and 0.72, respectively[1].

II. RELATED WORK

Mental illness profoundly affects an individual's emotions, cognitive processes, and social interactions. Its pervasive impact underscores the need for innovative prevention and intervention strategies within societies. Early detection of mental health issues is crucial to implementing these strategies effectively.[4]

Anu Priya et al [1], focused on predicting anxiety,depression and stress in modern life.This research focused on detecting anxiety,depression and stress using Depression, Anxiety and Stress Scale questionnaire(DASS21) .Data were collected from a total of 348 participants aged between 20 and 60 years via google form and subsequently classified using five machine learning algorithms -namely Decision Tree,Random Forest Tree,Naïve Bayes,Support Vector Machine and K-NN. Of the entire dataset Random Forest Tree was found to be highest with an average f1 score of 1.94.

Prince Kumar et al [2] , undertook assessment of anxiety , depression and stress using machine learning models. In total 39,776 instances were collected through online questionnaires between 2017 and 2019 by different methods using DASS21 and DASS42. Among different algorithms RBFN performed the best for depression in both the datasets. However, the result of Random Forest Tree is 100 percent for anxiety in DASS21.

Alanazi Rayan et al [3] ,focused on forecasting the mental well-being using machine learning. The total of 1065 posts from a Russian psychology form was used to construct a dataset. The characteristics of major depressive disorders and anxiety disorders were categorized using statistical and diagnostic Manual of mental disorders categories(DSM-5). The Area Under the Curve(AUC) showed that support vector machine(SVM) and Random Forest had the highest AUC's of 0.754(95 percent CI 0.700-0.800 and 95 percent CI 0.697-0.805 correspondingly).

III. METHODOLOGY

A. Data Collection

In total 1260 instances were collected through a survey dataset consisting of various features such as age, gender, self_employed, worklife, seek_life etc.,

Features	Description
Age	Age of Employee
Gender	Male or female
self_employed	Self employed or working in company
Family_history	Any one in family has mental problem
No_employees	Number of employees in company
Tech_company	IT company or not
Remote_work	Remote work or not
benefits	Receiving employee benefits or not

B. Classification

1. Logistic Regression:

Logistic Regression models the probability that a given input belongs to a particular class. It estimates the parameters of the logistic function using optimization algorithms such as gradient descent. Given a new data point, it calculates the probability that it belongs to each class using the logistic function and assigns it to the class with the highest probability.

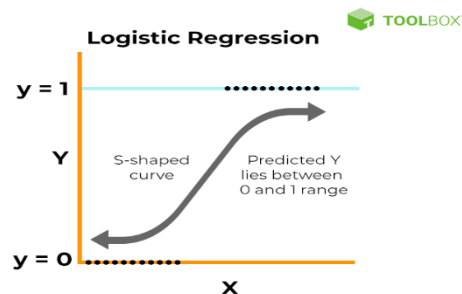


Fig 1. Logistic Regression

2. Decision Tree:

Decision tree recursively splits the data based on feature attributes to create a tree structure where each internal node represents a feature, each branch represents a decision based on that feature, and each leaf node represents a class label.

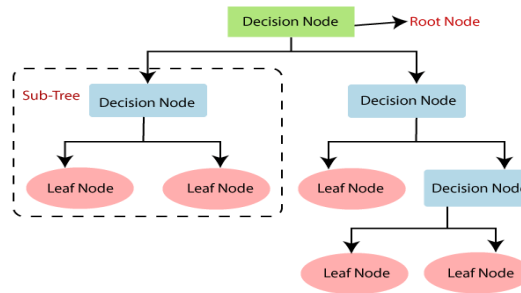


Fig 2. Decision Tree

3. Random Forest:

Random Forest is an ensemble method that builds multiple decision trees during training. Each tree is trained on a random subset of training data and a random subset of the features. For classification, each tree in the forest predicts the class, and the final prediction is made by taking a majority vote over all trees.

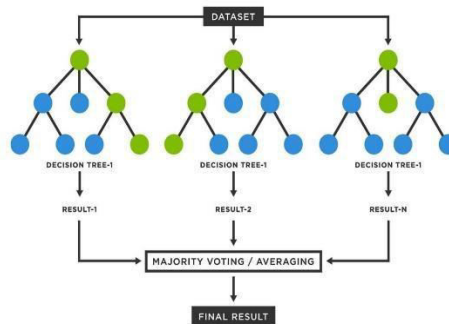


Fig 3. Random Forest

4. Naïve Bayes:

Naïve Bayes is a probabilistic classifier based on Bayes' theorem with the 'naïve' assumption of independence between features. It calculates the probabilities of each class given a set of input, and then selects the class with highest probability as the prediction.

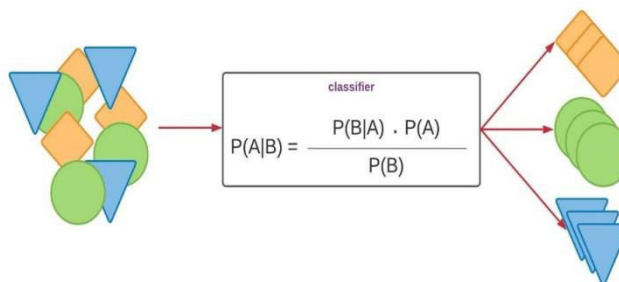


Fig 4. Naïve Bayes

5. Support Vector Machine:

SVM finds the hyperplane that best separates the classes in the feature space by maximizing the margin between the closest data points.

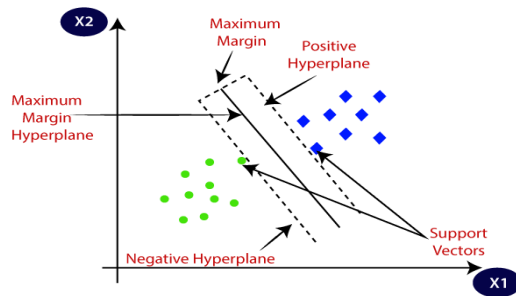


Fig 5. Support Vector Machine

6. *K-Nearest Neighbour:*

KNN stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions like Euclidean distance). When predicting the class of a new data point, KNN looks at the 'k' closest data points in the training set and assigns the majority class among them to the new point.

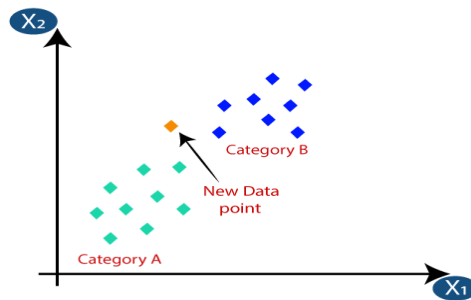


Fig 6. K-NN

IV. RESULTS

Six classification algorithms such as Logistic Regression, Decision Tree(DT), Random Forest, Naïve Bayes, Support Vector Machine(SVM) and K-Nearest Neighbour(K-NN) are applied to the dataset.

Equations from 1-6 are used to calculate Precision, Recall, Sensitivity, Specificity, F1 Score, Accuracy respectively.

$$Precision = TP/(TP+FP) \tag{1}$$

$$Recall = TP/(TP+FN) \tag{2}$$

$$Sensitivity = TP/(TP+FN) \tag{3}$$

$$Specificity = TN/(TN+FP) \tag{4}$$

$$F1\ Score = 2*(Precision*Recall)/(Precision+Recall) \tag{5}$$

$$Accuracy = TP/Total\ number\ of\ instances \tag{6}$$

Where,

TP (True Positive) = Diagonals of matrix

FN (False Negative) = Sum of consistent row for class

FP (False Positive) = Sum of corresponding column for class

TN (True Negative) = Sum of all rows and columns

Parameters	LR	DT	RF	SVM	KNN	NB
Precision	0.97	1	0.99	0	0.96	0.94
Recall	0.97	1	1	0	0.97	0.95
F1-Score	0.97	1	0.99	0	0.97	0.96
Sensitivity	0.97	1	1	0	0.97	0.98
Specificity	0.98	1	0.99	1	0.98	0.97
Accuracy (%)	98.38	100	99.59	70.56	97.98	97.58

TABLE I. PERFORMANCE MEASUREMENTS OF ML ALGORITHMS

Parameters	LR	DT	RF	SVM	KNN	NB
TP	173	173	173	175	172	170
TN	71	73	73	0	71	72
FP	2	0	0	73	2	1
FN	2	0	1	0	3	5
TPR	0.98	1	0.99	1	0.98	0.97
TNR	0.97	1	1	0	0.97	0.98
FPR	0.02	0	0	1	0.02	0.01
FNR	0.01	0	0	0	0.01	0.02
PPV	0.98	1	1	0.70	0.98	0.99
NPV	0.97	1	0.98	-	0.95	0.93

TABLE II. PERFORMANCE MEASUREMENTS OF CONFUSION MATRIX

V. CONCLUSION

This paper explored the potential of machine learning to predict employee mental health. While the results demonstrate the ability to flag at-risk individuals, it is crucial to recognize this as a tool for early intervention, not a definitive diagnosis. Further research is needed to improve accuracy and address ethical considerations around data collection. However, this paper highlights the potential of machine learning to play a significant role in promoting proactive mental health initiatives within the workplace.

REFERENCES

1. Anu Priya . (2019)" Predicting Anxiety, Depression and Strees in Modern Life using Machine Learning Algorithms."Procedia Computer Science 167 (2020) 1258-1267.
2. Prince Kumar . (2019)"Assessment of Anxiety, Depression and Stress using Machine Learning Models." Procedia Computer Science 171 (2020) 1989-1998.
3. Alanazi Rayan . (2023)"A novel approach to forecasting the mental well-being using machine learning."Alexandria Engineering Journal 84 (2023) 175-183.
4. Jetli Chung and Jason Teo . (2022)"Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges." Hindawi ID 9970363.
5. Konda Vaishnavi . (2022)" Predicting Mental Health Illness using Machine Learning Algorithms." Journal of Physics 2161(2022) 012021.
6. <https://adaa.org/understanding-anxiety/depression/symptoms>
7. https://www.webmd.com/balance/stress-management/stress-symptoms-effects_of-stress-on-the-body#1.
8. Saha, B., Nguyen, T., Phung, D., Venkatesh, S. (2016) "A framework for classifying online mental health-related communities with an interest in depression." *IEEE journal of biomedical and health informatics* **20** (4): 1008-1015.
9. Reece, A. G., Reagan, A. J., Lix, K. L. M., Dodds, P. S., Danforth, C. M., Langer, E. J.(2016) "Forecasting the Onset and Course of Mental Illness with Twitter Data." *Scientific reports* **7** (1): 13006.
10. Saitta, L., (2000) "Support-Vector Networks." *Kluwer Acad. Publ. Bost.* : 273–297.

11. Hamed, T., Dara, R., Kremer, S. C. (2014) "An accurate, fast embedded feature selection for SVMs." *In 2014 13th International Conference on Machine Learning and Applications* IEEE :135-140.
12. Martinez-Arroyo, M., Sucar, L. E. (2006) "Learning an optimal naive bayes classifier." *In 18th International Conference on Pattern Recognition (ICPR'06)* IEEE 3: 1236-1239.
13. Cheng, J., Greiner, R. (1999) "Comparing Bayesian network classifiers." *In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* Morgan Kaufmann Publishers Inc : 101-108.
14. Taneja, S., Gupta, C., Goyal, K., Gureja, D. (2014) "An enhanced k-nearest neighbor algorithm using information gain and clustering." *Fourth International Conference on Advanced Computing & Communication Technologies* IEEE : 325-329.
15. V. Rocha-Rego, J. Jogia, A. F. Marquand, J. Mourao-Miranda, A. Simmons, and S. Frangou, "Examination of the predictive value of structural magnetic resonance scans in bipolar disorder: a pattern classification approach," *Psychological Medicine*, vol. 44, no. 3, pp. 519–532, 2013.
16. D. Grotegerd, T. Suslow, J. Bauer et al., "Discriminating unipolar and bipolar depression by means of fMRI and pattern classification: a pilot study," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 263, no. 2, pp. 119–131, 2012.
17. G. Valenza, M. Nardelli, A. Lanata et al., "Predicting mood changes in bipolar disorder through heartbeat nonlinear dynamics," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1034–1043, 2016.
18. J. Mourão-Miranda, J. R. Almeida, S. Hassel et al., "Pattern recognition analyses of brain activation elicited by happy and neutral faces in unipolar and bipolar depression," *Bipolar Disorders*, vol. 14, no. 4, pp. 451–460, 2012.
19. G. Roberts, A. Lord, A. Frankland et al., "Functional dysconnection of the inferior frontal gyrus in young people with bipolar disorder or at genetic high risk," *Biological Psychiatry*, vol. 81, no. 8, pp. 718–727, 2017.
20. G. Akinci, E. Polat, and O. M. Koçak, "A video based eye detection system for bipolar disorder diagnosis," in *Proceedings of the 2012 20th Signal Processing and Communications Applications Conference (SIU)*, Mugla, Turkey, April 2012.
21. M.-J. Wu, I. C. Passos, I. E. Bauer et al., "Individualized identification of euthymic bipolar disorder using the cambridge neuropsychological test automated battery (CANTAB) and machine learning," *Journal of Affective Disorders*, vol. 192, pp. 219–225, 2016.
22. A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with twitter data," *Scientific Reports*, vol. 7, no. 1, 2017.
23. D. Leightley, V. Williamson, J. Darby, and N. T. Fear,
24. "Identifying probable post-traumatic stress disorder: applying supervised machine learning to data from a UK military cohort," *Journal of Mental Health*, vol. 28, no. 1, pp. 34–41, 2018.
25. S. Papini, D. Pisner, J. Shumake et al., "Ensemble machine learning prediction of posttraumatic stress disorder screening status after emergency room hospitalization," *Journal of Anxiety Disorders*, vol. 60, pp. 35–42, 2018.
26. D. Conrad, S. Wilker, A. Pfeiffer et al., "Does trauma event type matter in the assessment of traumatic load?" *European Journal of Psychotraumatology*, vol. 8, no. 1, Article ID 1344079, 2017.
27. C. R. Marmar, A. D. Brown, M. Qian et al., "Speech-based markers for posttraumatic stress disorder in US veterans," *Depression and Anxiety*, vol. 36, no. 7, pp. 607–616, 2019.
28. D. Vergyri, B. Knoth, E. Shriberg et al., "Speech-based assessment of PTSD in a military population using diverse feature classes," in *Proceedings of the INTERSPEECH 2015*, Dresden, Germany, 2015.
29. L. E. Salminen, R. A. Morey, B. C. Riedel, N. Jahanshad, E. L. Dennis, and P. M. Thompson, "Adaptive identification of cortical and subcortical imaging markers of early life stress and posttraumatic stress disorder," *Journal of Neuroimaging*, vol. 29, no. 3, pp. 335–343, 2019.
30. D. Rangaprakash, G. Deshpande, T. A. Daniel et al., "Compromised hippocampus-striatum pathway as a potential imaging biomarker of mild-traumatic brain injury and posttraumatic stress disorder," *Human Brain Mapping*, vol. 38, no. 6, pp. 2843–2864, 2017.
31. M. Sumathi and B. Poorna, "Prediction of mental health problems among children using machine learning techniques," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 1, 2016.
32. A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, "Predicting mental health problems in adolescence using machine learning techniques," *PLoS One*, vol. 15, no. 4, Article ID e0230389, 2020.
33. A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLoS One*, vol. 14, no. 11, Article ID e0224365, 2019.
34. S. K. Ernala, M. L. Birnbaum, K. A. Candan et al., "Methodological gaps in predicting mental health states from social media," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, UK, May 2019.
35. Sarah Graham, Colin Depp, Ellen E Lee, Camille Nebeker, Xin Tu, Ho-Cheol Kim, and Dilip V Jeste. Artificial intelligence for mental health and mental illnesses: an overview. *Current psychiatry reports*, 21(11):1–18, 2019.
36. Theodoros Iliou, Georgia Konstantopoulou, Mandani Ntekouli, Christina Lymperopoulou, Konstantinos Assimakopoulos, Dimitrios Galiatsatos, and George Anastassopoulos. Iliou machine learning preprocessing method for depression type prediction. *Evolving Systems*, 10(1):29–39, 2019.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details