

A Survey on Use of Machine Learning for Employee Wellness

Chetan Thote, Prof. Jareena Shaikh

Department of Computer Engineering, Zeal College of Engineering & Research, Pune, Maharashtra, India

ABSTRACT

The Pandemic times that we are living in has forced many changes into the lifestyle and preferences of our lives. Mental health is the most talked about topic in recent times with cases of employee burnout happening frequently. And because of such burnouts at workplace Employee churn rate is also high in such cases. In this paper, survey of the work done in the field of machine learning to predict stress, anxiety and depression is presented. The survey paper is providing analysis and extensive review of the work presented around the topic by more than 10 papers in last decade. This survey paper discusses different approaches taken by various researchers around the topic, it also talks about machine learning algorithms used by these researchers to create machine learning model.

Keywords – Machine Learning, Mental health, Survey.

Article Info

Volume9, Issue 3

Page Number: 589-593

Publication Issue

May-June-2022

Article History

Accepted :01May2022

Published :12 May 2022

I. INTRODUCTION

In post covid era, Mental health has been the talk of most of the discussions that are happening around the world. Various research papers, surveys, articles have been published around the topic of mental health awareness. This survey paper talks about the mental health of employees at workplace. Working is good for mental health but working at an organization where employee wellness is not looked into can lead to mental and physical health problems. Depression and anxiety have a very significant economic impact as well on global economy as lost productivity cost around US\$ 1 trillion per year [1]. Research papers at various conferences around the world have talked about developing machine learning algorithms to help understand impact of different variables on predicting

depression, stress and anxiety around humans. Depression can lead to memory loss, lack of concentration, an inability to make decisions, overeating and weight gain. Similarly, Stress can lead one to feeling upset, decrease in energy levels, headaches and overreaction. Identifying such symptoms early on using machine learning can not only help employees by managing their workload, it can also help organizations improve productivity of employees and growth of organization.

II. LITERATURE SURVEY

The impact of various variables such as breathing difficulty, Heart rate, nervousness and many such variables of our daily life in predicting stress, anxiety, Depression, dissatisfaction and frustration using

machine learning algorithms was discussed by Anu Priya et al. in [2]. The use of five different algorithms Decision Tree, Random Forest, Naïve Bayes, Support Vector Machine and K – Nearest Neighbour were studied to predict five levels of Anxiety, depression and stress, on a dataset of around 350 participants aged between 20 and 60 years, both males and females, employed and unemployed individuals through Depression, Anxiety and Stress Scale questionnaire. The questionnaire comprises of 21 questions, with 7 questions allocated to each of the scales of Depression, Anxiety and Stress. Following data collection, responses were encoded using numeric values of 0 to 3, and the scores were then calculated by adding the values associated with each question. Once the final scores had been calculated, these were labelled according to severity shown in table 1. The findings of the study show that, the variable ‘Scared without any reason’ was most important for Anxiety scale; the variable ‘Life was meaningless’ was found to be important on Depression scale whilst ‘Difficult to relax’ was important on Stress Scale. Naïve Bayes was found to be more accurate but the Random Forest classifier was identified to be the best model due to the fact that this problem produced imbalanced classes.

TABLE I Table 1: Severity levels

Category	Anxiety	Depression	Stress
Normal	0-7	0-9	0-14
Mild	8-9	10-13	15-18
Moderate	10-14	14-20	19-25
Severe	15-19	21-27	26-33
Extremely Severe	20+	28+	33+

In a similar research done by Arkaprabha Sau and Ishita Bhakta in 2018[3] to compare the performance of machine learning algorithms to identify anxiety and depression among seafarers. Around 500 seafarers were interviewed at a Dock Complex in India for information related to their health, demography and occupations. The presence of Anxiety and depression

was assessed by the Hospital Anxiety and Depression Scale. The status of anxiety and depression during interview were assessed by HAM-A and HAM-D [8,9]. As per HAM A and HAM D score, a person could be labelled either “No Anxiety - No Depression” or “No Anxiety - Depression” or “Anxiety - Depression” or “Anxiety - Depression”. This parameter/variable was by as a dependant variable for machine learning algorithms. Recursive feature elimination method was used to identify predictor features. Machine learning classifiers such as CatBoost, Logical Regression, Naïve Bayes, Random Forest, Support Vector Machine were evaluated to assess the presence of anxiety and depression in seafarers and it was observed that CatBoost was the most satisfactory measure for this research, with accuracy of 82.6%, precision of 84.1% and ROC area 0.882. ROC curve of Five classifiers are shown in Fig 1.

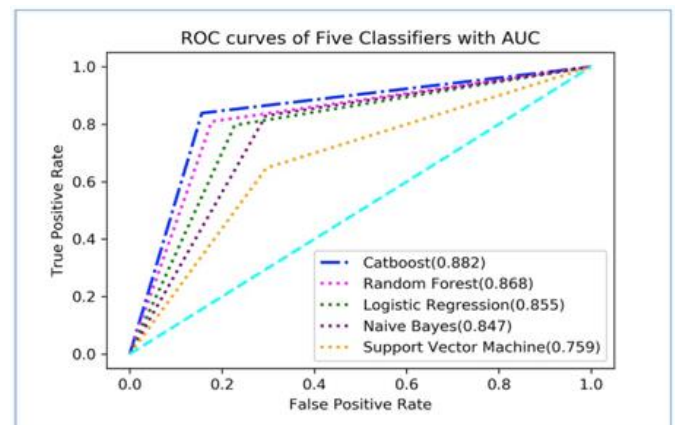


Fig. 1. ROC curves of 5 classifiers on training set with 10 fold cross validation

Anusha Garlapati et al in 2020 worked on Predicting employees under stress for pre-emptive remediation using Machine learning algorithms. During this research they have analysed the consequence of stress on employee appearance. Moreover, this influences physical ailments and a lack of commitment to work. Through this work, they intend to analyse the stress level that employees are subjected to owing to a phenomenon like the present pandemic. Here, machine learning algorithms are used to predict whether employees undergoing stress or not. Data set

used for this research belonged to a set of workers where parameters like ID, Age, Average Daily hours and so on were captured. The proportion of the target variable in the data is of symmetry 4.67:1. Data was standardized before applying algorithms as the data consists of categorical values also. Numeric data is done by using Standardization whereas categorical data is done by using Label Encoder. It ought to only standardize numeric data and should not standardize or normalize categorical data. XG Boost classifier was studied as XG Boost has tremendous predictive power and is about 10 times more durable than other gradient boosting techniques. It holds a variety of regularization which diminishes overfitting and enhances overall performance. Consequently, it is further recognized as the “regularized boosting” technique. Like it has true positive, true negative, false positive, false negative values. Used to evaluate the performance of the classification model.

Isaac Teoh Yi Zhe & Pantea Keikhosrokiani in 2021 studied use of Deep learning models to predict mental workload of workers [5]. Knowledge workers face a high mental workload in terms of planning and coordination. One solution is to predict the mental workload of knowledge workers. Some machine learning models have been implemented for mental workload prediction, but deep learning models are yet to be introduced for this purpose. Deep learning models are superior to machine learning models because of their ability to correct inaccurate predictions if they ever occur. Therefore, this study aims to optimize the extreme learning adaptive neuro-fuzzy inference system (ELANFIS) by integrating particle swarm optimization into a micro-genetic algorithm to predict the mental workload of knowledge workers. Although the adaptive neuro-fuzzy inference system (ANFIS) shows reasonable prediction performance, it also suffers from the curse of dimensionality and has a poor computation time. Thus, ELANFIS is introduced because its curse of dimensionality is less severe when solving problems with a high number of input dimensions. The

integration of the advantages of a micro-genetic algorithm and particle swarm optimization is suggested to optimize the premise parameters of ELANFIS, as this can allow better solutions to be located at a faster rate. The proposed model yields promising prediction results, with improvements of 6.0665 in the Mean Squared Error (MSE) and 1.279 in the Root Mean Squared Error (RMSE) for regression; the proposed model even surpasses the prediction results of ELANFIS optimized with PSO alone, with improvements of 1.5369 in MSE and 0.4094 in RMSE for regression. The findings are expected to assist employers in determining an appropriate working lifestyle for their employees.

Abhisek Sethy and Dr. Ajit Kumar Raut, studied attrition rate of employee in an organization based on the turnover [1]. Machine learning algorithms can assist firms in predicting staff attrition. Analysts can construct and prepare a machine learning model that can anticipate the personnel who are leaving the firm using previous data kept in human resources (HR) departments. The data of around 1500 employees with 32 variables was acquired using IBM Watson Analytics. Five machine learning models (Regression Analysis, Bayes Classifier, SVM, KNN, RFT) were studied, and it was observed that effort – reward imbalance is most likely fundamental general explanation for attrition.

III. RESULTS & COMPARISON

Prediction of stress in employees is a classification problem. And various classification algorithms are being used by various researchers to predict stress levels of employees. It is observed that Random forest and XGBoost are the widely used classification methods to solve this problem. Both algorithms offer the better accuracy as compare to other algorithms. Deep learning models are also very useful for prediction. It also shows the maximum output but it takes more time as compared to other algorithms. We also found out that variables such as Age, Gender,

Average Daily hours, Demography, occupation, and behavioural pattern plays important role in developing such models. Details of the research paper studied in this survey are listed in Table 2.

IV. SUMMARY

The importance of correlating performance of employees with mental health of employees along with the occupations, socio-demographics and various other factors was discussed in this paper. Data were collected using standard questionnaire around the performance and mental health of employee by various researchers. Classification techniques were used such as Decision Tree, Random Forest, Support Vector Machine (SVM), Naïve Bayes, K- Nearest Neighbour (K-NN).

V. ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor Prof. Ajinkya Dawale, without whom I would not take up the research in the first place. I would like to take this opportunity to thank my guide Prof. Jareena Shaikh and Head of the Department Prof. A. V. Mote for their valuable guidance and for providing their valuable insights on my work.

TABLE II Comparison

Authors	Problem	Data Mining Tech	Recommend
Anu Priya, Shruti Garga, Neha Prerna Tigga	Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms	Decision Tree, Random Forest, Naive Bayes, Support Vector Machine, K-NN	Random Forest

Arka prabha Sau and Ishita Bhakta	Screening of anxiety and depression among seafarers using machine learning technology	CatBoost, Logistic Regression, Naïve Bayes, Random Forest, and Support Vector Machine	CatBoost
Anusha G, Dorendra Radha K, Kavya G, Gayathri N	Predicting Employees under Stress using Machine learning Algorithm	XGB classifier	XGB classifier
Rohit Punnoose, Pankaj Ajit	Prediction of Employee Turnover in Organizations using Machine Learning Algorithms	XGBoost, Logistic Regression, Naïve Bayesian, Random Forest (Depth controlled), SVM(RBF kernel), LDA, KNN (Euclidean distance)	XGBoost

VI. REFERENCES

- [1]. <https://www.who.int/teams/mental-health-and-substance-use/promotion-prevention/mental-health-in-the-workplace>
- [2]. Neha Prerna Tigga, Shruti Garga, Anu Priya, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms" ICCIDS 2019

- [3]. Arkaprabha Sau, Ishita Bhakta (2018) "Screening of anxiety and depression among seafarers using machine learning technology" Informatics in Medicine Unlocked, Volume 16, 2019, Pages 100228.
- [4]. A. Sethy, Dr. Ajit Kumar Raut , "Employee attrition rate prediction using machine learning approach", Turkish Journal of Physiotherapy and Rehabilitation; 32(3)
- [5]. Isaac Teoh Yi Zhe & Pantea Keikhosrokiani. "Knowledge workers mental workload prediction using optimised ELANFIS", Applied Intelligence volume 51, pages2406–2430 (2021)
- [6]. Tyshchenko, Y. (2018)"Depression and anxiety detection from blog posts data."Nature Precis. Sci., Inst. Comput. Sci., Univ. Tartu, Tartu,Estonia.
- [7]. Shear MK, Vander Bilt J, Rucci P, Endicott J, Lydiard B, Otto MW, Pollack MH, Chandler L, Williams J, Ali A, Frank DM. Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH-A). *Depress Anxiety* 2001 Jan 1;13(4):166–78.
- [8]. Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. *Arch Gen Psychiatr* 1988 Aug 1;45(8):742–7.
- [9]. Pankaj Ajit, Rohit Punnoose , " Prediction of Employee Turnover in Organizations using Machine Learning Algorithms ", International Journal of Advanced Research in Artificial Intelligence, 2016
- [10]. Oei, T. P., Sawang, S., Goh, Y. W., Mukhtar, F. (2013) "Using the depression anxiety stress scale 21 (DASS-21) across cultures." *International Journal of Psychology* 48 (6): 1018-1029.
- [11]. S. Jahan, "Human Resources Information System : A Theoretical Perspective", *Journal of Human Resource and Sustainability Studies*, 2014.
- [12]. E. K. Kalokerinos, J. D. Henry and C. von Hippel, "Stereotype threat among older employees: Relationship with job attitudes and turnover intentions", *Psychology and aging*, 2013.
- [13]. Li, L., Zhang, X. (2010) "Study of data mining algorithm based on decision tree." *International Conference On Computer Design and Applications IEEE* 1: V1-155.
- [14]. Paul, A., Mukherjee, D. P., Das, P., Gangopadhyay, A., Chintla, A. R., Kundu, S. (2018) "Improved random forest for classification." *IEEE Transactions on Image Processing*
- [15]. Marjorie Laura Kane-Sellers , "To explore various personal, as well as work variables impacting employee voluntary turnover " 2009
- [16]. S. J. Delany and P. Cunningham, "k-Nearest neighbour classifiers", *Multiple Classifier Systems*, 1- 17, 2007.
- [17]. A. Liaw and M. Wiener, "Classification and regression by randomForest", *R news*, 2002.
- [18]. L. Breiman, *Random forests*. Machine Learning, 2001.
- [19]. V. Vapnik and C. Cortes, *Support-vector networks*. Machine learning, 1995.