
ANALYZING MENTAL HEALTH BETWEEN VARIOUS PROFESSIONS USING MACHINE LEARNING AND NLP MODEL

¹T. Abirami, G.K. Kamalam, ²T. Madhuvanathi, S. Akshita, Y. Annuncia Marena, and M. Haritha

¹Department of Information Technology, Kongu Engineering College, Perundurai, India.

²Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, India.

Email: abi.it@kongu.ac.in; madhuvanathi313@gmail.com

Abstract

A mental illness is a mental health condition that come to possession in the way of thinking, relating to others, and day-to-day routine. Abundance of mental illnesses have been identified and elucidated. Society, work pressure, family, and other factors may all play a vital role in why a person is in a particular circumstance. The main concept of this research is to forecast such illnesses by determining what the person is going through mentally by collecting data containing questionnaires through survey by focusing mainly two professions such as IT jobs and Teaching (Professors). Target groups used for this process are HOD, Associate professor, Assistant Professor, Professor, HR, Manager, TL, AE. Many individuals who are working in IT are affecting mentally due to stress and some other factors. Nowadays, the teaching profession is also become one of the toughest fields to handle, which involves both physical and mental efforts in dealing with work and students. So, using Logistic Regression, Decision Tree, K- Nearest Neighbor, Random Forest, and Natural language processing algorithms for creating ensemble models and further comparing the models. Gathered data from professors to get accurate statistics. The highest accuracy that obtained is 83% from decision Tree by processing IT professionals and 96% from Decision Tree for real-time data by processing teaching professionals.

Keywords. Mental Health, Machine Learning, Natural Language Process, IT Employees, Professors, Real-Time data.

1. INTRODUCTION

In this human era, the impact of mental health plays a major role in today's lifestyle. A poor state of mental health leads to various side effects which destroys the entire life of a common individual.[6] Many factors are involved in affecting one's mental health like work environment, family, society, overthinking, traumas etc. Now-a-days people are stressed out in the work place due to peer pressure, work load, lack of confidence. Occupational ill health, poor productivity, and human mistake are all linked to work-related stress. This translates to more sick days, high staff turnover, and poor organizational effectiveness, as well as an increase in accidents owing to human error. Workplace stress can cause heart disease, back pain, headaches, gastrointestinal disorders, or a variety of minor ailments, as well as psychological impacts including anxiety and despair, lack of attention, and poor decision-making. [4] This project is main focus on analysing the impact of mental health issues for working individuals. People in every profession[1] will experience a stress/anxiety in certain point.

Here for the analysis [2] two set of people who is working in two different streams which is IT profession and teaching profession are considered. Currently these two fields are becoming more demand and the growth of technology as made it more competitive. So, there may be a high chance for people to be affected by stress and unstable mental health.

Currently people are more affected in balancing both physical and mental health because of making things smooth in personal and work life. Monitoring and maintaining the mental health help a person to lead a happy life. [7] Thus, an analysis for mental condition of the working people is made by implementing Random Forest, Decision Tree, KNN and logical regression in machine learning and NLP and predicted the accuracy. Also made an visualization for easy understanding between the attributes which has more impact on the analysis.

1.1. Literature Survey

This [5]research paper has examined the state of mental health in a target group using Decision Trees, SVM, Naive Bayes Classifier, Logistic Regression, and KNN Classifier. [3] A support vector machine and Naive-Bayes classifier is implemented in the study. This paper aims to use natural language processing on Twitter feeds to do depressive emotion analysis. The key contribution is the creation of an experimental paradigm for successfully creating stress at numerous levels, as well as a framework for identifying stress at various levels using EEG data analysis. The [11] employed algorithms are Decision Trees (DT), K-NN, Nave Bayes (NB), Random Forest Trees (RFT), and Support Vector Machines in the study (SVM). The Depression, Anxiety, and Stress Scale questionnaire was used to collect data from working and unemployed people from many cultures and communities [15]. The main goal of this systematic review was to summarise and characterise these strategies in methodological and technical terms, as well as to examine the potential utility in mental health clinical practise. The employed Machine Learning, Depression, Anxiety, Support Vector Machines, Convolutional Neural Networks, Linear Regression, K Nearest Neighbor Classifiers, and Linear Discriminant Analysis in this study. The goal of [7] study is to present a system that uses machine learning algorithms to identify depression and anxiety disorders. To determine the best algorithm for detecting the severity of anxiety or depression, the performance metrics of each algorithm were evaluated. Support vector machines, decision trees, naive bayes classifier, K-nearest neighbor classifier, and logistic regression were employed in [12] study. The purpose of this study is to propose using Machine Learning algorithms to identify the condition of mental health in a target population such as high school kids, college students, and working professionals. Random Forest Classifier, Random Forest Regression, Naive Bayes, and K Neighbor Classifier were examined in [9]study. The purpose of this study is to find out which sources of stress are associated with stress-related symptoms in people by looking at age, monthly income, gender, occupation, children, city, prior employment, marital status, and current job satisfaction level.

2. METHODOLOGY

To compare and predict the mental health issues between the Tech people and the professors using different machine learning algorithms and Natural language processing. A real time dataset has been collected by taking survey from the professors in the respected institution and also from nearby colleges for teaching professionals. For IT professionals the dataset has been taken from Kaggle. In both the data has been collected by asking some questionnaires about the work environment, family history and mental condition. Here the target attribute is taken has treatment and predicted the accuracy.

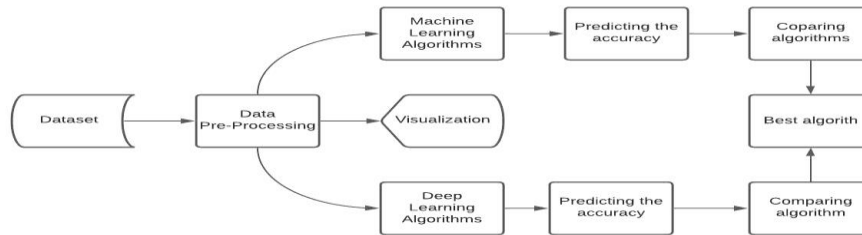


Fig.1.Process Flow Diagram

The above figure, 1 shows the process flow of the project. Initially the dataset is taken from the Kaggle and data pre-processing has been done by using python then after that implemented the ML and DL technique and obtained the accuracy. Finally graphical representation of the mental health data has been performed along with the performance analysis.

2.1. Data Pre-processing

For real-time data acquired from teaching professionals and data collected from individuals working in the IT area, data pre-processing is carried out using Python. It entails locating missing values and substituting mean values instead of it. Also, feature selection was used to suit the data. The data is then divided into train and test models for ease of algorithm implementation. The degree of variables in the dataset should be discovered and computed, as this information is useful for better data preparation to match the expectations of machine learning algorithms.

2.2. Implementation of ML and NLP Algorithm for IT profession

Using the accuracy _score, Precision, Random Forest Regressor, linear Regressor, AUC Score, Cross-validated AUC from the sklearn package. Following the import of these libraries, chosen the appropriate features like family history, work culture, work balance etc., from the data to perform the algorithms and obtain accurate results. Improved the accuracy of the existing machine learning algorithm using logistic regression, from 79 % to 83.05 %, and implemented other algorithms such as Decision tree regression (83.05%), Random Forest (81.25%), and k Nearest Neighbor (81.25%). Among this the decision tree, which maps an observation item to its target value, is the most accurate. Leaves indicate categories, and branches reflect feature combinations that lead to those classifications in these tree topologies. The decision tree is straightforward to grasp and analyses, and data preparation is minimal. Otherwise, it is capable of reliably handling both numerical and categorical data. For the same Mental Health data, a deep learning Natural Language Processing (NLP) algorithm was used, with an accuracy of 85%. Finally, a performance analysis graph has been implemented by taking, the X-axis as the ML algorithms and Y-axis as the accuracy which has been predicted. This helps for better understanding.

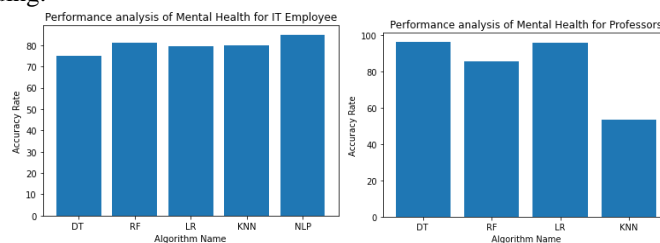


Fig.2: Analysis of IT Employee and Professors

2.3. Implementing ML Algorithm for Teaching Professors

Using the libraries accuracy score, mean squared error, Random Forest Regressor, svm, linear Regressor, and r2 score from the sklearn package. Following the import of these libraries, choose the appropriate features from the data to implement the algorithms for correct results here target variable(x) is Age, Gender, Family_history, Benefits, Care_options, Anonymity, Leave and work_interface then defendant variable(y) also called as output is feature_cols. Professors data was pre-processed in the same way, by finding mean values and substitute it with the missing values. After that, the data is separated into trained and tested categories. The machine learning methods were then implemented, and the accuracy was attained. Logistic regression (96 %), K-Nearest Neighbor (53.57 %), decision tree (96.42 %), and random forest are the ML algorithms that were implemented and predicted the accuracy (85.71 %) The decision tree has the highest accuracy here as well.

2.4. Experimental Analysis and Result

The accuracy analysis of IT employee and professor is demonstrated using the proposed method. Thus, acquired a high degree accuracy of 96% using real time data that is collected from the college by linear regression and other algorithms such as Decision tree, Random Forest, K- Nearest Neighbor, and Natural Language Processing for employees and attributes with age, gender, mental health issue at work, work experience, and role.

2.5. Data Visualization

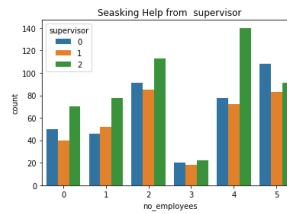
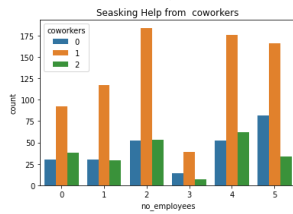


Fig.3: Seeking Help from Co-workers Fig.4:Seeking Help from Supervisor

From the above figure 3&4, the bar graph is implemented between the total number of counts and the number of employees for seeking help from co-workers and supervisor [8] Here the inputs are maybe, yes, no which are received from the IT professionals for seeking help from the co-workers for getting support for mental health issues. That is converted into 0, 1, and 2 for easy processing using python. Thus the output shows that the preference is YES by most of the people

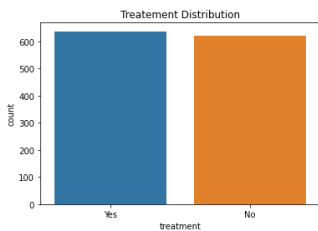


Fig.5: Treatment Distribution

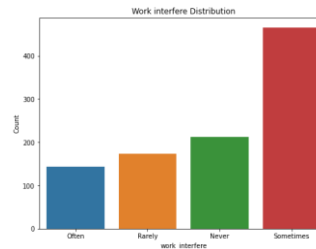


Fig.6: Work Interfere Distribution

From the above fig 5, shows the number of people who prefers to take treatment for mental health impacts like stress, depression, anxiety. The attributes here taken for x-axis is treatment and total count for y-axis. The result shows that highest percent preference is yes. This has made to understand that maximum number of people is affected mentally from the work.From figure 6, the x-index indicates the

work_interfere rate like often, rarely, never, sometimes and y-index indicates the total count. The graph shows that most of the IT people has answered sometimes for interfering mental health with the work life.

3. CONCLUSION

As depression, stress, anxiety which all comes under the mental health issues are recognized as a severe burden on healthcare systems all around the world. The key challenge is predicting whether or not someone has a mental illness. With one-on-one physical encounters, psychologists can examine and treat the patients. For IT employees and professors, the proposed idea is for analysing mental health conditions. K-Nearest Neighbor, Decision Tree, Random Forest, and Natural Language Processing were utilized to process those data. The most accurate method is Decision Tree, which had a 96 percent accuracy rate for IT professionals and 83 percent for Professors. And also did a graphical illustration of the stress prediction analysis for the two occupations. Machine learning and natural language processing methods may provide numerous viewpoints in research studies, but it should also be viewed as clinical practice support tools. Availability of more medical related data for mentally affected persons may help the future work with more accuracy rate. As a result, doubling the volume of treatment can lower one's stress levels.

4. FUTURE SCOPE

Mental health affects everyone in different ways, but haven't always been allowed to talk about it at work. Not just these two professionals are at the biggest risk of generating mental health problems. People of all ages are being influenced mentally in today's generation as a result of technological advancements and behaviour modification, whether they are aware of it or not. In the future, this will aid in the analysis and prediction of a more accurate cause for this illness by taking into account other attributes such as affected people's behaviour that differs from that of a healthy person, what measures people are taking to overcome this issue, how a mentally impacted person should be treated, and so on. Employers, for the most part, play an important role in supporting employees' mental health. This might take the shape of benefits, but it also means ensuring that staffs have the resources they need to complete their tasks and avoid burnout.

5. REFERENCES

- [1] Ahmed, A., Sultana, R., Ullas, M. T. R., Begom, M., Rahi, M. M. I., & Alam, M. A. (2020). *A machine learning approach to detect depression and anxiety using supervised learning*. Paper presented at the 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). DOI: 10.1109/CSDE50874.2020.9411642.
- [2] Cellini, P., Pigoni, A., Delvecchio, G., Moltrasio, C., & Brambilla, P. (2022). *Machine learning in the prediction of postpartum depression: A review*. Journal of affective disorders. DOI: 10.1016/j.jad.2022.04.093.
- [3] Deshpande, M., & Rao, V. (2017). *Depression detection using emotion artificial intelligence*. Paper presented at the 2017 international conference on intelligent sustainable systems (iciss). DOI: 10.1109/ISS1.2017.8389299.
- [4] Kene, A., & Thakare, S. (2021). *Mental Stress Level Prediction and Classification based on Machine Learning*. Paper presented at the 2021 Smart Technologies, Communication and Robotics (STCR). DOI: 10.1109/STCR51658.2021.9588803.

- [5] Le Glaz, A., Haralambous, Y., Kim-Dufor, D.-H., Lenca, P., Billot, R., Ryan, T. C., . . . Berrouiguet, S. (2021). *Machine learning and natural language processing in mental health: Systematic review*. Journal of Medical Internet Research, 23(5), e15708. DOI: 10.2196/15708.
- [6] Liu, Y. S., Chokka, S., Cao, B., & Chokka, P. R. (2021). *Screening for bipolar disorder in a tertiary mental health centre using EarlyDetect: A machine learning-based pilot study*. Journal of Affective Disorders Reports, 6, 100215. DOI: <https://doi.org/10.1016/j.jadr.2021.100215>.
- [7] Lu, H., Uddin, S., Hajati, F., Khushi, M., & Moni, M. A. (2022). *Predictive risk modelling in mental health issues using machine learning on graphs* Australasian Computer Science Week 2022 (pp. 168-175). DOI: <https://doi.org/10.1145/3511616.3513112>.
- [8] Moon, N. N., Mariam, A., Sharmin, S., Islam, M. M., Nur, F. N., & Debnath, N. (2021). *Machine learning approach to predict the depression in job sectors in Bangladesh*. Current Research in Behavioral Sciences, 2, 100058. DOI: <https://doi.org/10.1016/j.crbeha.2021.100058>.
- [9] Narayanrao, P. V., & Kumari, P. L. S. (2020). *Analysis of machine learning algorithms for predicting depression*. Paper presented at the 2020 international conference on computer science, engineering and applications (iccsea). DOI: 10.1109/ICCSEA49143.2020.9132963.
- [10] Park, H., & Lee, K. (2022). *Prediction of suicidal ideation in shift workers compared to non-shift workers using machine learning techniques*. Journal of affective disorders, 307, 125-132. DOI <https://doi.org/10.1016/j.jad.2022.03.076>.
- [11] Priya, A., Garg, S., & Tigga, N. P. (2020). *Predicting anxiety, depression and stress in modern life using machine learning algorithms*. Procedia Computer Science, 167, 1258-1267. DOI: <https://doi.org/10.1016/j.procs.2020.03.442>.
- [12] Reddy, U. S., Thota, A. V., & Dharun, A. (2018). *Machine learning techniques for stress prediction in working employees*. Paper presented at the 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). DOI: 10.1109/ICCIC.2018.8782395.
- [13] Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). *Behavioral modeling for mental health using machine learning algorithms*. Journal of medical systems, 42(5), 1-12. DOI:10.1007/s10916-018-0934-5.
- [14] R.Bharathi, T.Abirami,” Energy efficient compressive sensing with predictive model for IoT based medical data transmission”, Journal of Ambient Intelligence and Humanized Computing, November 2020, <https://doi.org/10.1007/s12652-020-02670-z>
- [14] Qiao,J, A Systematic Review of Machine Learning Approaches for Mental Disorder Prediction on Social Media. 2020 International Conference on Computing and Data Science (CDS), DOI: 10.1109/CDS49703.2020.00091