Prediction of Mental Disorder for employees in IT Industry

Sandhya P, Mahek Kantesaria

Abstract: Mental health is nowadays a topic which is most frequently discussed when it comes to research but least frequently discussed when it comes to the personal life. The well-being of a person is the measure of mental health. The increasing use of technology will lead to a lifestyle of less physical work. Also, the constant pressure on an employee in any industry will make more vulnerable to mental disorder. These vulnerabilities consist of peer pressure, anxiety attack, depression, and many more. Here we have taken the dataset of the questionnaires which were asked to an IT industry employee. Based on their answers the result is derived. Here output will be that the person needs an attention or not. Different machine learning techniques are used to get the results. This prediction also tells us that it is very important for an IT employee to get the regular mental health check up to tract their health. The employers should have a medical service provided in their company and they should also give benefits for the affected employees.

Index Terms— mental health in IT, machine learning in mental health, machine learning, mental health

I. INTRODUCTION

Mental health is the aggregation of emotional, social and psychological well-being of a person. It effects on the person's thinking, acting and feeling capability. Mental health is a measure of handling stress and decision making with every step in life. Mental Health is very important factor in every stage in life whether it be childhood or an adult. Mostly mental health is something which never discussed publicly, and no proper awareness is there in society. People would generally not talk about it in public. Mental health could affect one's thinking and behavior. Some common reasons of instable mental health could be:

- Past life experiences, such as ragging or bullying
- Biological factors, such as genes
- Hereditary problem from family

Mental health problems are very common if people accepts it and they can get better over it. If someone is experiencing few below mentioned symptoms then they might have mental health problem:

- Too less or too much of sleep
- Going away from usual routine and people
- Sudden increase in intake of drugs
- Severe mood swings
- Thinking to harm yourself

Positive mental health will support person to have full potential. It will also him to cope up with the stress at home and workplace. It will increase productivity of the people. Now to maintain this positive mental health one should get help from others, connect with others, help others, have usual routine and develop coping skills.

II. LITERATURE SURVEY:

Inappropriate diagnosis of the mental health could lead to incorrect treatment. This could in long run prove to be as fatal as premature death. Hence this paper introduces a new approach as semi-automated system. This system will diagnose psychological disorder in patient. This is achieved by comparing patient's mental health with the DSM-IV-TR, Fourth edition revision. Here they are using genetic algorithm, classification and machine learning techniques to build semi-automated system. The final future goal is to fully automate the system unless it meets the required standards of identifying patient's mental illness. The classifier had an assessment of a patient as accurate as possible. The final call will be of analyst for treatment of patient.

People with mental disorder often face other anxiety disorder which eventually develops into depression. Hence authors are interested in online communities for data. They have crawled data from 247 online communities of 80,000 users. Then they have extracted the psycho-linguistic posts based on topics, which served as input to model. Machine learning techniques are applied to generate joint model for identifying mental health related features. At last they performed empirical validation of model on dataset where model performs best in recent techniques.

With outburst of social media usage, health scientists have become more active, particularly for data related to disease epidemics and analyze warning signals of mental health issues. In this paper also, they have focused on precursor, cognitive distortion and symptoms of psychological behavior like anorexia, anxiety and depression. They have gathered personal blogs from Tumblr API and labeled them whether they show distorted behavior or not. Then they apply LIWC (Linguistic Inquiry and Word Count) to extract features and further machine learning is applied further. They found out that it is actually possible to detect distortions by system from personal blogs with accuracy of 73.0% and false negative rate of 30.4%.

According to studies, stress will tend to increase depression, heart attack, stroke and cardiac arrest. To measure stress level which could improve the harmful effects associated with health. In this paper machine learning along with electroencephalogram (EEG) signal analysis is used. In this experiment stress is added to experiment based on Montreal Imaging Stress Task (MIST). The machine learning with EEG gives feature extraction, selection, classification (logistic regression, support vector

Revised Manuscript Received on April 03, 2019.
  Sandhya P, Associate Professor SCSE Vellore Institute of Technology, Chennai Campus
  Mahek Kantesaria, M.Tech. CSE SCSE Vellore Institute of Technology, Chennai Campus

machine and naive bayes) and 10 fold cross validation. The result with proposed framework is accuracy of 94.6% for two-level stress identification and accuracy is 83.4% for multi-level stress identification. So the proposed method framework tends outperform than required.

Depression is such a disease which not affect mental health but also physical health. There are studies which prove depression and type-s diabetes go hand in hand. The authors talk about the same issue with supervised machine learning technique. They develop a model to allocate class label with the help of features to dataset. Classification helps to associate class labels to subjects with known feature matrix and unknown class labels. This paper proves that classification of machine learning techniques is proving to be encouraging for detection of depression and type 2 diabetes among patients.

Machine learning and artificial intelligence can be applied to betterment of the society. The authors have proposed many machine learning techniques like support vector machine, naives bayes, decision tree, K-nearest neighbor and logistic regression. Here the targeted group of people are students and working professionals. There is a questionnaire which is asked to these people. Then unsupervised techniques were applied, and mean opinion score is calculated.

Symptoms of mental disorder could be easily observed on social media platform. These are being automatically able to locate by methods. Here the authors studied about social media activity to detect depression and other mental illness. Mentally ill people are distinguished by their membership in online forums, online screening or community distribution analysis on Twitter. These users are detected by their regularity in online presence and their use of languages. Many methods could be applied to detect mentally ill people on social media.

Early detection of mental illness could be useful to doctors to start the medication for patient. Here authors have use one of the deep learning technique called as Convolutional Neural Network (CNN). This uses AlexNet and build end to end CNN. The input to this system is combination of Landmark Motion History Image, Motion History Image and Gabor Motion History Image which generated depression dataset. Accuracy of all these techniques combined is 83%.

This research have managed to have identify five mental health problem by help of eight machine learning methods. Dataset has 60 cases. From whole dataset 25 features are identified as important. Feature selection algorithm is applied on full attribute dataset. Accuracy of selected attribute set and full attribute set is compared. Classifiers like multiclass classifier, LAD tree and multilayer perceptron generate more accurate result. Only slight performance difference is observed between them.

Emotional intelligence is a new emerging field in text mining. The advent of social media allows to have ample of data required for sentiment analysis of images and text. Here authors apply natural language processing on Twitter posts for having emotional conclusion for depression. Individual tweets are given labels as negative or neutral based on words used and their closeness to depression relevancy. Class prediction is done by Support Vector Machine and Naïve-Bayes classifier. The results are in F1-score, confusion matrix and accuracy.

III. IMPLEMENTATION:

Here the dataset used is a survey taken among the IT professionals from different regions. It mainly has information like age, gender, location of work, type of work, is he/she self-employed? and many more. Then now for machine learning to apply, we fir need to remove unnecessary fields like comments and timestamp. This step is called as data cleaning. Data cleaning is the process by which data gets rid off from not required data, making it appropriate for further analysis. Incorrect format, errors while capturing, missing data acts as garbage data. Many of the attributes have empty values as input so default values will be assigned to it. For Integer it is 0, float is 0.0 and string is NaN. Now for gender attribute we have make it all in standard format by replacing any unknown inputs to standard input. The next step is encoding the data. Now after data is again checked whether nay data is missing or not. Then the dataset is scaled and fitted.

Now machine learning techniques are applied and compared that which suits the dataset best. The first algorithm is logistic regression. Logistic regression is a sigmoid function having an S-shaped curve that takes any real value and maps to value from 0 to 1. The equation is $y = e^{(b0 + b1*x)} / (1 + e^{(b0 + b1*x)})$. The second algorithm is K Nearest Neighbors classifier. Prediction for new instance is made by training dataset and finding similar K instances. The similarity is checked distance between new instance and other instances. The most common is Euclidean distance of formula $EuclideanDistance(x, xi) = sqrt(sum((xj - xij)^2))$. The others are Hamming distance, Manhattan distance and Minkowski distance. Third algorithm is decision tree. In decision tree classification is method where data is split to certain parameter. The leaves are called as final results. Nodes are called decision points where data is split. Entropy and information gain are two important features calculated in decision tree classification. Random forest is the fourth technique. Random forest is an ensemble learning method that works by constructing a multitude of decision trees in training phase and output is mean or mode of classes. Fifth technique is bagging and boosting. Bagging is Bootstrap procedure with high variance of decision tree. Bagging doesn't overfit trees. Hence single decision tress are grown very deep and they are not pruned. Boosting is procedure where it uses weighted averages to make stronger learners from weak learners. Boosting works in team unlike bagging which works on individual tree. Last very technique is Neural Network. It modeled same as human brain to recognize patterns. This method will interpret sensory data and label and cluster them.

IV. RESULT & EVALUATION:

Here while applying these techniques to dataset we also evaluate methods by following metrics:
- Classification Accuracy: it is a measure of correct

predictions in percentage
- Confusion Matrix: Tabular format to describe performance of model

  o True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN)
- Precision: it measures how often correct positive value is predicted.
- AUC Score: it is area under ROC (Receiver Operating Characteristic) curve in percentage.

Model	Accuracy (%)	Confusion Matrix				Precision (%)	AUC Score (%)	Cross-Validated AUC (%)
		TP	TN	FP	FN			
Logistic Regression	79.9	161	141	50	26	76.3	79.9	87.5
Knn	80.4	169	135	56	18	75.1	80.5	87.8
Decision Tree	80.6	175	130	61	12	74.1	80.8	88
Random Forest	81.2	174	133	58	13	75	81.3	89.3
Bagging	77.7	143	151	40	44	78.1	77.7	84.3
Boosting	81.7	172	137	54	15	76.1	81.8	87.4
Neural Network	81.4	-	-	-	-	75.5	87.6	84.5

Here more the table we can see that Random forest is the best model trained for the dataset. The important features required are Gender, Age, Country, self_employed and family_history. A csv file is prepared with Id number as first column and predicted column is mental treatment required to patient as second column.
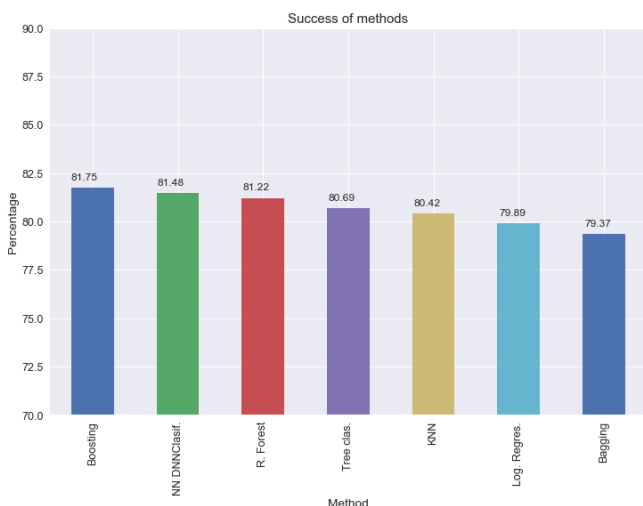

Fig. 1- Accuracy of all methods

V. CONCLUSION:

There are many suggestions that employers and employees could keep in mind. Employers need to keep track of number of their employees having mental disorder. Employers should allow flexible work environment with flexible work scheduling and break timings. They should allow employees to work from home or have flexible place of work. They should give day-to-day feedback and guidance for nurturing employees' health. This type of model could be used to detect metal health progress among employees and also could lead to policy changes. Employees could talk to colleagues and their managers about their problem freely. Hence upper management could help them to get correct aid with beneficiaries like work from home, flexible timings, more leaves, many more. Employees should know health benefits provided by their organization participate in any wellness programs. Proper feedback should be provided to employee when they resign from their job. This could help them to improve their health.

VI. FUTURE WORK:

At every interval of months, the survey should be taken in IT companies to know how well their employees are. It should be a rule that should be embedded in their policy book. This project could further be developed with beautification and could potentially be attached to stack of default software in an organization.

REFERENCES:

1. G. Azar, C. Gloster, N. El-Bathy, S. Yu, R. H. Neela and I. Alothman, "Intelligent data mining and machine learning for mental health diagnosis using genetic algorithm," *2015 IEEE International Conference on Electro/Information Technology (EIT)*, Dekalb, IL, 2015, pp. 201-206.
2. B. Saha, T. Nguyen, D. Phung and S. Venkatesh, "A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression," in *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1008-1015, July 2016.
3. T. Simms, C. Ramstedt, M. Rich, M. Richards, T. Martinez and C. Giraud-Carrier, "Detecting Cognitive Distortions Through Machine Learning Text Analytics," *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, Park City, UT, 2017, pp. 508-512.
4. A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel and A. S. Malik, "Machine Learning Framework for the Detection of Mental Stress at Multiple Levels," in *IEEE Access*, vol. 5, pp. 13545-13556, 2017.
5. R. M. Khalil and A. Al-Jumaily, "Machine learning based prediction of depression among type 2 diabetic patients," *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nanjing, 2017, pp. 1-5.
6. Srividya, M & Subramaniam, Mohanavalli & Natarajan, Bhalaji. (2018). Behavioral Modeling for Mental Health using Machine Learning Algorithms. Journal of Medical Systems. 42. 88. 10.1007/s10916-018-0934-5.
7. Ashar Khan, Mohd Shahid Husain & Anam Khan. (2018). Analysis of Mental State of Users using Social Media to predict Depression! A Survey. International Journal of Advanced Research in Computer Science,
   II International Conference on "Advancement in Computer Engineering and Information Technology" Volume 9, Special Issue No. 2, pp. 100-106.
8. Megat S'adan, Megat Ahmad Haziq & Pampouchidou, Anastasia & Meriaudeau, Fabrice. (2018). Deep Learning Techniques for Depression Assessment. 10.1109/ICIAS.2018.8540634.
9. Sumathi, Ms & B, Dr. (2016). Prediction of Mental Health Problems Among Children Using Machine Learning Techniques. International Journal of Advanced Computer Science and Applications. 7. 10.14569/IJACSA.2016.070176.
10. M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, Palladam, 2017, pp. 858-862.