CLASSIFICATION ALGORITHM BASED MENTAL HEALTH PREDICTION USING DATA MINING

[1]Mrs. M. Kavitha [2]Dr. Madhavi Pingili [3]M. Spurthi, [4]SK. Fayaz, [5]K. Kirthan, [6]S. Santhosh, ,
[1]Assistant Professor, Dept. of IT, CMR Engineering College, UGC Autonomous, Kandlakoya, Medchal Road, Medchal Dist, Hyderabad-501 401,e-mail : kavitha.madas@gmail.com
[2]Associate Professor and HOD, Dept. of IT, CMR Engineering College, UGC Autonomous, Kandlakoya, Medchal Road, Medchal Dist, Hyderabad-501 401
[3,4,5,6]B.Tech Scholars, Dept. of IT, CMR Engineering College, UGC Autonomous, Kandlakoya, Medchal Road, Medchal Dist, Hyderabad-501 401

ABSTRACT: The emotional, psychological and social welfare of a person is revealed by their mental health. It influences how an individual will think, feel or handle a situation. Positive mental health helps an individual to work productively and achieve their full potential. At each point in life, mental health is vital, from childhood to adulthood. Numerous factors contribute to mental health issues which lead to mental illness like stress, social anxiety, depression, obsessive compulsive disorder, drug addiction, workplace issues and personality disorders. The onset of mental illness should be determined without flaws for maintaining an appropriate life balance. We have collected data from online available datasets. The data has been label encoded for better prediction. The data is being subject to various machine learning techniques to obtain labels. These classified labels will then be used to build a model to predict the mental health of an individual. The accuracy of the algorithm will be analyzed before it is used to build the model. We planned to implement classification algorithms such as Decision Tree, Random Forest and Naïve Bayes. Our target population is in the working class i.e. people above the age of 18. Once the model is built, it will be integrated to a website so that it can predict the outcome as per the details provided by the user.

Keywords: Random Forest, Decision Tree, like stress, social anxiety, depression, obsessive compulsive disorder, drug addiction, workplace issues and personality disorders.

I. INTRODUCTION

The emotional, psychological and social welfare of a person is revealed by their mental health. It influences how an individual will think, feel or handle a situation. Positive mental health helps an individual to work productively and achieve their full potential. At each point in life, mental health is vital, from childhood to adulthood. Numerous factors contribute to mental health issues which lead to mental illness like stress, social anxiety, depression, obsessive compulsive disorder, drug addiction, workplace issues and personality disorders. The onset of mental illness should be determined without flaws for maintaining an appropriate life balance. Mental health prediction is one of the most essential parts of reducing the probability of serious mental illness. Meanwhile, mental health prediction can provide a theoretical basis for public health department to work out psychological intervention plans for medical workers.

Currently, some representative scales are usually used to measure mental health, such as Selfreporting Inventory (SCL-90) Minnesota Multiphasic Personality Inventory (MMPI), Self-Rating Anxiety Scale (SAS), Self-Rating Depression Scale (SDS), Eysenck Personality Questionnaire (EPQ), the Sixteen Personality Factor Questionnaire (16PF). Above scales are widely used internationally because they are guided by various psychological theories and can transform abstract mental health concepts into observable specific indicators. However, some shortcomings are not considered in the scales mentioned above. First, the different emphasis of scale measurement leads to the differences in the

evaluation criteria because many factors need to be considered in the measurement of mental status. Second, the existing way of answering the scales is self-evaluation, which inevitably makes the respondent hold something back. Third, a lot of time is spent in obtaining the results of the scale for judging mental status in emergency situations. Although the diagnosis and intervention of mental symptoms are significant, prevention is even more important. Therefore, using existing information to predict mental health is of great significance.

The evaluation of mental wellness is extremely critical to understand and suggest therapies for patients with a deviated mental behavior. Most individuals are prone to stress while some are affected by depression due to various reasons. An administrative panel of World Health Organization (WHO) assessed in 2011 that, by 2030, depression will be the chief source of worldwide disease burden. There is a fundamental change to incorporate the mental fitness outline of an affected individual by healthcare providers and it will be made obligatory in the approaching years to deliver improved medication and also promote speedy recoveries.

For our project, we have collected data from online available dataset, provided by an OSMI (Open Sourcing Mental Illness) survey. The dataset mainly consists of data of working individuals. It will predominantly be beneficial for employers and employees by creating greater awareness about work related mental illness. We have applied machine learning algorithm to create a model. It has been implemented on a website for users to get knowledge about their mental illness. The webpage shows a probability and recommendation to the user based on the inputs provided.

II. LITERATURE SURVEY

U. S. Reddy, A. V. Thota and A. Dharun et. al [1] Machine Learning Techniques for Stress Prediction in Working Employees. Various Machine Learning techniques were used. The accuracy of the models was obtained and studied comparatively. Boosting had the highest accuracy among the models implemented. By using Decision Trees, prominent features that influence stress were identified as gender, family history and availability of health benefits in the workplace.

M. P. Dooshima, E. N. Chidozie, B. J. Ademola, O. O. Sekoni, I. P. Adebayo et. al [2], A Predictive Model for the Risk of Mental Illness in Nigeria Using Data Mining. Naïve Bayes' and the Decision Trees' Classifiers were used to formulate the predictive model for the risk of mental illness based on the identified and validated variables using the WEKA software. Decision Trees algorithm can assist mental health experts to apply the rules deduced by the algorithm for the early detection of mental illness.

M. Srividya, M. Subramaniam and B. Natarajan et. al [3] Behavioral Modeling for Mental Health using Machine Learning Algorithms. It proposes to apply various machine learning algorithms such as support vector machines, decision trees, naïve bayes classifier, K-nearest neighbor classifier and logistic regression to identify state of mental health in a target group.

S. G. Alonso, I. Torre-Díez, S. Hamrioui, M.l López-Coronado, D. C. Barreno, L. M. Nozaleda, and M. Franco et.al [4] It can be

said that use of Data Mining techniques applied to diseases such as dementia, schizophrenia, depression, etc. can be of great help to the clinical decision, diagnosis prediction and improve the patient's quality of life.

M. A. Haziq Megat S'adan, A. Pampouchidou and F. Meriaudeau, et. al [5] Deep Learning Techniques for Depression Assessment assessment of depression is done by applying three deep learning techniques of Convolutional Neural Network (CNN). These techniques are transfer learning using AlexNet, fine-tuning using AlexNet and building an end-to-end CNN. The inputs of the CNNs are a combination of Motion History Image, Landmark Motion History Image and Gabor Motion History Image, and have been generated on a depression dataset. Accuracy of the three deep learning techniques is computed. As of now, transfer learning technique achieved a result comparable to the state of the art, of 83% accuracy.

D.Filip & C. Jesus et.al [6] A Neural Network Based Model for Predicting Psychological Conditions. It presents a neural network model that is able to predict the likelihood of developing psychological conditions such as anxiety, behavioral disorders, depression, and post-traumatic stress disorder.

Deziel M, Olawo D, Truchon L. & Golab et. al [7] Analyzing the Mental Health of Engineering Students using Classification and regression created a survey based on Canadian guidelines and applied classification and regression algorithms to collected data. The results reveal interesting relationships between various aspects of mental health and year of study.

Tomar, D., & Agarwal, S. et. al [8] A survey on Data Mining approaches for Health care. International Journal of Bio-Science and Biotechnology. Explores the utility of var ious Data Mining techniques such as classification, clustering, association, regression in health domain. In this paper, we present a brief introduction of these techniques and their advantages and disadvantages and also highlights applications, challenges and future issues of Data Mining in healthcare. Recommendation regarding the suitable choice of available Data Mining technique is also discussed here.

III. PROPOSED METHOD

Based on the above survey, we have designed a system with the primary goal of developing a website where users can enter values in a form and get results about potential or current mental illness based on their input. Firstly, we have collected a dataset which is available online. The data gathered is analyzed and pre-processed. The data contains different labels such as age, gender, distance of workplace from home, previous mental illness, family history etc. We have label encoded the data for better prediction. We implemented the Decision Tree and Random Forest algorithms for testing the data and to find the more accurate algorithm. We have applied Decision Tree classification algorithm for classification of the data as it was found to be more accurate. We analyzed the data with the help of this algorithm to find various insights that the data revealed. We then created a model based on the decision tree algorithm and used it on the website we designed.

According to our goal, we have designed a website where a user shall login and fill up a form 4 which has questions based on the dataset gathered. The user will answer the

questions and a result about his/her mental condition will be provided on the website as per the inputs provided. The website makes use of the model that we have built using the machine learning algorithms to provide the output. Since this project makes use of a dataset related to workplace mental illness, it will help raise awareness among employees and employers to provide greater attention to work related stress, depression and proper benefits can be provided to employees suffering from a mental illness.

IV. SOFTWARE DESIGN

4.1 System Architecture:

The methodology of review consists following steps

1. Data Collection
2. Feature Extraction
3. Classification
4. Testing
5. URL feature prediction

4.1.1 Data Collection: The dataset was composed of URLs of twitter data set along with their respective identity such as malicious or non-malicious. Based upon these data the given input can be classified as either malicious or non-malicious. The dataset which we use in this project is the twitter data set collected.

4.1.2 Feature Extraction: The feature extraction technique plays an important role. The features are the main parameter that are involved for classification of URL. Texture extraction is determined as the example of information or course of action of the structure with random interval.

4.1.3 Classification: In a typical classification system image captured by a camera and then processed. In Supervised classification, most importantly preparing occurred through known gathering of pixels. The numbers of clusters decided by users. When trained pixels are not available, the supervised classification is used that is KNN.

4.1.4 Testing: In the testing phase the URLs are being tested.

4.1.5 URL feature prediction: Finally, we get the URL is either malicious or non-malicious.

4.2 Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It may be used as a communication tool between a system analyst and any person who plays a part in the order that acts as a starting point for redesigning a system. The DFD is also called as a data flow graph or bubble chart.

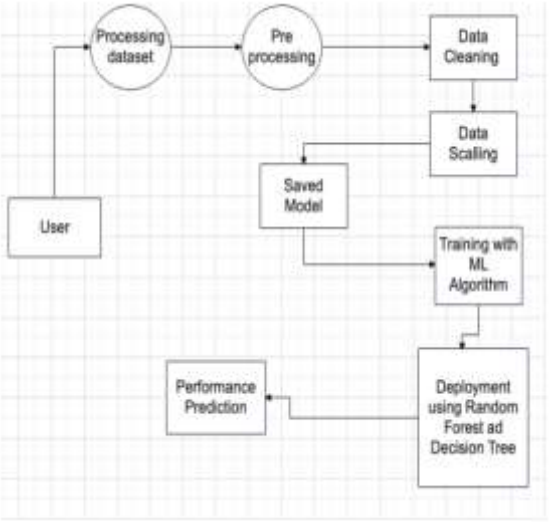

Fig.1: System Architecture

4.3 UML Diagrams:

UML is a standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems. UML was created by the Object Management Group (OMG) and UML 1.0 specification draft was

proposed to the OMG in January 1997. There are several types of UML diagrams and each one of them serves a different purpose regardless of whether it is being designed before the implementation or after (as part of documentation). UML has a direct relation with object-oriented analysis and design. After some standardization, UML has become an OMG standard. The two broadest categories that encompass all other types are:
1. Behavioural UML diagram and
2. Structural UML diagram.

As the name suggests, some UML diagrams try to analyse and depict the structure of a system or process, whereas others describe the behaviour of the system, its actors, and its building components. The different types are broken down as follows:
1. Sequence diagram
2. Use case diagram
3. Activity diagram
4. Class diagram

4.3.1 Sequence diagram: A sequence diagram simply depicts interaction between objects in a sequential order i.e., the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.

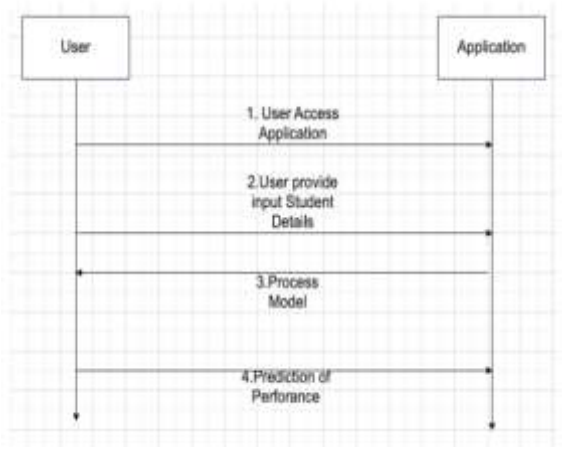

Fig.2: Sequence diagram

4.3.2 Use case diagram: A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use case in which the user is involved. A use case diagram is used to structure the behaviour thing in a model. The use cases are represented by either circles or ellipses. User needs to give the data. Then the System will give the results. System Consist of data set that will pre-process the data and then split the data and apply the train and test the data then it will predict the results

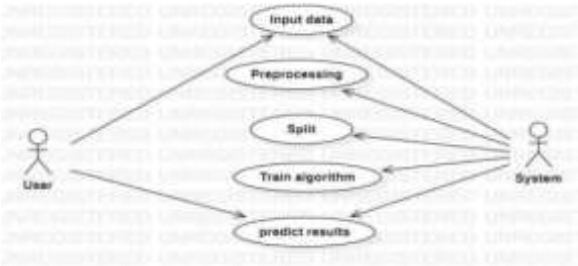

Fig.3: Use case diagram

Use case: User needs to give the data. Then the System will give the results. System Consist of data set that will pre-process the data and then split the data and apply the train and test the data then it will predict the results.

4.3.3 Activity diagram: Activity diagram is another important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all types of flow control by using different elements such as fork, join, etc.

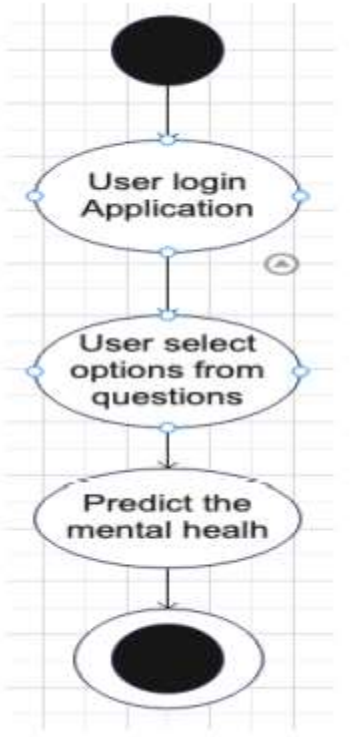

Fig.4: Activity diagram

4.3.4 Class diagram: In software engineering, a class diagram in the Unified Modelling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

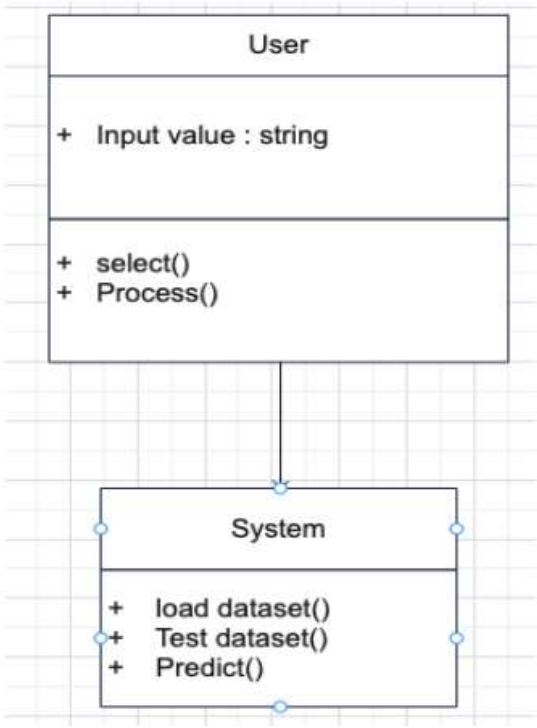

Fig.5: Class diagram

V. OUTPUT SCREENS

Fig.6: Input screen 1

A list of questions is asked to an individual about his feelings like if he feels tired for no reason or if he he/she feels nervous for no reason etc. The inputs are recorded and used for mental health prediction of an individual.

Fig.7: Input screen 2

Fig.8: Input screen 3

After successfully recording the inputs, the system will predict the mental health of an individual if he is suffering from mental illness or not. If he is not suffering from mental illness the system suggests the individual not to consult a counsellor to treat his mental condition.

Fig.9: Output screen

If he is suffering from mental illness the system suggests the individual to consult a counsellor to treat his mental condition.

Fig.10: Output screen

VI. CONCLUSION & FUTURE SCOPE

Mental health is an extremely sensitive and important topic currently. It is integral for living a healthy and balanced life. Mental health impacts one's thoughts, behaviour and emotions. It can affect the productivity and effectiveness of an individual. As per the study by WHO, depression will be a major cause of mental illness in the world and people need to take more care about their mental wellbeing for a balanced social and professional life. People who are hesitant to approach humans for diagnosis can make use of online predictors for results. To do the prediction, we have encoded the data first.

We have then used the decision tree algorithm and trained a model which we have used on our website. The accuracy we received with decision tree was 82% with 258 instances of data being classified correctly out of 315 instances. When the user answers the questions on our webpage, he/she gets a probability of their mental health condition as well as recommendations. Due to the accuracy, we achieved, it can be concluded that the output displays the correct result and the chance of the illness being misclassified is minimal. In the future, we can create a system which predicts a specific mental illness that a

person suffers from, however extensive data collection needs to be carried out for it.

FUTURE SCOPE:

By using classification based mental health prediction using data mining project in future, the mental health of an individual can be predicted and also the system will be able to suggest the selfcare tips to improve his mental health.

VII. REFERENCES

[1]. U. S. Reddy, A. V. Thota and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 2018, pp. 1-4.

[2]. M. P. Dooshima, E. N. Chidozie, B. J. Ademola, O. O. Sekoni, I. P. Adebayo, "A Predictive Model for the Risk of Mental Illness in Nigeria Using Data Mining", International Journal of Immunology. Vol. 6, No. 1, 2018, pp. 5-16.

[3] M. Srividya, M. Subramaniam and B. Natarajan, "Behavioral Modeling for Mental Health using Machine Learning Algorithms", Journal of Medical Systems Vol. 42(5):88 May 2018

[4] S. G. Alonso, I. Torre-Díez, S. Hamrioui, M.l López-Coronado, D. C. Barreno, L. M. Nozaleda, and M. Franco. "Data Mining Algorithms and Techniques in Mental Health: A Systematic Review", Journal of Medical Systems 42(9), July 2018

[5] M. A. Haziq Megat S'adan, A. Pampouchidou and F. Meriaudeau, "Deep Learning Techniques for Depression Assessment," 2018 International Conference on Intelligent and Advanced System (ICIAS), Kuala Lumpur.

[6] D. Filip & C. Jesus. (2015). "A Neural Network Based Model for Predicting Psychological Conditions", International Conference on Brain Informatics and Health 252-261

[7] Deziel, M, Olawo, D., Truchon, L., & Golab, L. "Analyzing the Mental Health of Engineering Students using Classification and Regression". Published in EDM 6 July 2013.

[8] Tomar, D, & Agarwal, S. (2013). "A survey on Data Mining approaches for Health care", International Journal of Bio-Science and Biotechnology.