PREDICTION OF MENTAL HEALTH USING MACHINE LEARNING

Dr.J.Arokia Renjit , Adlin Sajeesha M.J , Sangavai V.D , Sree Devi D.S
Head of the computer science and engineering department , Student , Student , Student
Computer Science and Engineering , Jeppiaar Engineering College , Chennai, India

Abstract - *Mental health problems detected at earlier stage helps the psychologist to medicate and enhance the patient's life. Depression is one of the leading causes of disability worldwide. This article provides an evaluation of machine learning in mental healthcare. The analysis of research on ML -peculiar to psychological health and a talks of how ML can support health while considering its present bounds concerning ML technology . Hence , there is an immediate requirement to cure basic psychological health problems that abound among children which may lead to serious issues, when not cured at an early stage. Machine learning Techniques are presently applicable for evaluating medical data and treating the problem. The attributes have been decreased by applying Feature Selection algorithms over the complete data set of attributes. The accuracy of all the algorithms are compared and the best algorithm with higher efficiency is been deployed using flask. However, alert is required to ignore unnecessary results, and more work is required to viaduct the gap between ML in psychological health research and health care.*

Keywords - Supervised learning ,logistic regression , random forest , naive bayes , decision tree classifier, flask deployment .

1. INTRODUCTION

Machine learning is to presage the future from past data. Machine learning (ML) provides computers the capability to learn without programming specifically . It aims on the development of Computer Programs which gets changed when revealed to new data . The process of instructing and predicting precludes the use of particular algorithms. It feeds the trained set of data to an algorithm, and that algorithm applies to give predictions on new set of test data. Machine learning can be divided into three divisions. There are supervised learning with both input and output, unsupervised learning with just input, and reinforcement learning. A supervised learning program is both given the input and the label to learn data has to be labeled by a human before itself. Unsupervised learning includes no labels. This algorithm figures out the grouping of input data. Finally, Reinforcement learning firmly interrelate with its environment and it regains positive or

negative feedback to enhance its performance. Data scientists use many kinds of machine learning algorithms to invent various patterns in python which led to insights which could be implemented . At a top most level, these algorithms can be divided into two groups based on the way they "learn" and predict the data.

This paper advances our research work in classifier-based prediction:

- Classification is the process of forecasting the class of given data points in which classes are represented as targets .

- Classification predictive modelling is the mission of generating a mapping function from input variables to provide output variables.

- In machine learning, classification is a supervised learning approach in which the computer program gets from the data input provide to it and then utilizes this to classify new scrutiny.

- This data set may generally be bi-class or multi-class . Some examples of classification problems are recognizing the speech, recognizing the handwriting, identifying the biometrics, classifying the documents, etc.
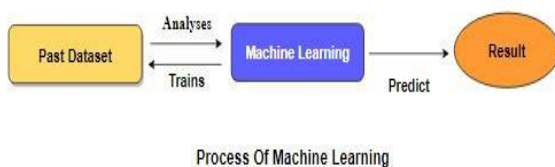

Process Of Machine Learning

figure 1.1

Supervised Machine Learning is the superiority practical machine learning which uses supervised learning. Supervised learning is one which includes

input and the output variables. This utilizes an algorithm to grasp the plotting function from the input to output is $y = f(X)$. The aim is to approximate the plotting function so that when you have new input data (X) ,it is easier to detect the output variables (y) for that data as mentioned in figure 1.1. Strategies of Supervised Machine Learning algorithms include **logistic regression**, **Decision Tree** and **support vector machine.** The data which is used to instruct and train the algorithm has been marked already with accurate answers. Supervised learning problems can be divided into **Classification** problems. This problem has as its aim to design a continuous model which could detect the value of the dependent attribute from the variables of attributes. The variation between the two chores is the actuality that the dependent attributes is either numerical or categorical. A classification model endeavours to design some results from observed values. When one or more inputs is given, a classification model will try to detect the value of one or more results.

2. LITERATURE SURVEY

A summary is a recap of important information about the source, but a synthesis is a re-organization, reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent or relevant of them.

2.1 Machine Learning-based Approach for Depression Detection in Twitter Using Content and Activity Features by Hatoon AlSagri, and Mourad Ykhlef mentioned that they aim to exploit machine learning techniques for detecting a probable depressed

Twitter user based on both network behaviour and tweets. For this purpose, they trained and tested classifiers to distinguish whether a user is depressed or not using features extracted from his/her activities in the network and tweets. The results showed that the more features are used, the higher the accuracy and F-measure scores in detecting depressed users. This method is a data-driven, predictive approach for the early detection of depression or other mental illnesses.

2.2 Predicting the Utilization of Mental Health Treatment with Various Machine Learning Algorithms by Meera Sharma

Sonak Mahapatra, Adeethyia Shankar mentioned that proper diagnosis and treatment for people with mental health disorders remains underdeveloped in modern-day's society due to the widely ever-present public stigma attached to caring about mental health. Recently there have been attempts in the data science world to predict if a person is suicidal (and other diagnostic approaches) yet all face major setbacks. To begin, big data has many ethical issues related to privacy and reusability without permission—especially in regards to using feeds from social media. Additionally, people diagnosed with specific mental health conditions may not actually seek treatment, so data may be incorrect. In this research, we address both of these problems by using anonymous datasets to predict the answer to a different question—whether or not people are seeking mental health treatment. We also use a large variety of machine learning and deep learning classifiers and predictive models to predict

with a high accuracy rate through statistical analysis. As a result, these individuals would be more productive, reducing social and economic costs in the tech workplace.

2.3 Prediction of Mental Disorder for employees in the IT Industry by Sandhya P,

Mahek Kantesaria mentioned that they've taken the dataset of the questionnaires which were asked to an IT industry employee. Based on their answers the result is derived. Here the output will be whether the person needs attention or not. Different machine learning techniques are used to get the results. This prediction also tells us that it is very important for IT employees to get a regular mental health check-up to track their health. The employers should have a medical service provided in their company and they should also give benefits to the affected employees There are many suggestions that employers and employees could keep in mind. Employers need to keep track of the number of their employees having mental disorders. Employers should allow a flexible work environment with flexible work scheduling and break timings. They should allow employees to work from home or have a flexible place of work.

2.4 Prediction of Mental Health Problems Among Children Using Machine Learning Techniques by Ms. Sumathi M.R, Dr. B. Poorna mentioned that twenty-five attributes have been identified as necessary for analysing the issues from the reports. The attributes have been decreased by applying

Feature Selection algorithms over the complete data set of attribute . The efficiency over the complete attribute set and selected attribute set on various machine learning techniques have been analysed. It is obvious from the final results that the three classifiers Multilayer Perceptron, Multiclass Classifier, and LAD Tree bring more efficient results and there is only a slight variation between their performances over complete attribute set and specific set of attribute . The research has compared eight machine learning techniques (classifiers) on classifying the dataset to different mental health problems. The data set is very minimal and in the future, the research may be applied to a large data set to obtain more accuracy.

3. IMPLEMENTATION

We have proposed an idea in which we collected datasets and pre processed the data for our project and mapped the data into graphs to check the quality of the data then train the data with different algorithms to predict the output with higher accuracy.
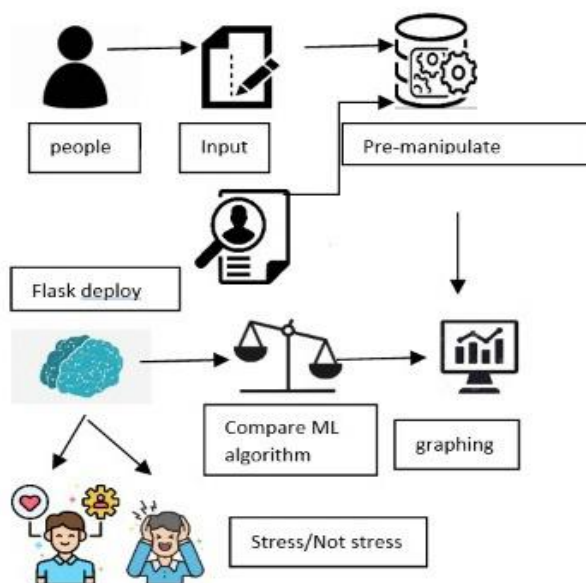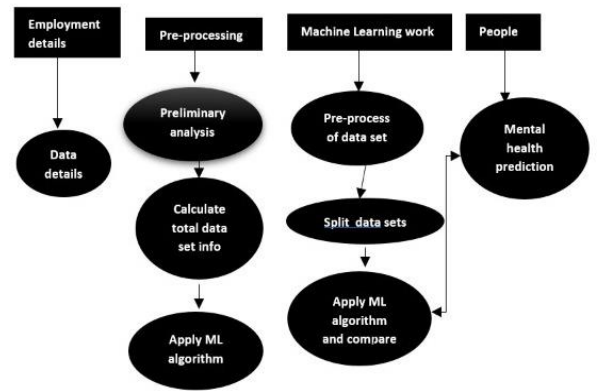

figure 3.1 System Architecture

figure 3.2 Activity Diagram

4. PROPOSED SYSTEM

The main scope of our project is to detect mental health Prediction and to design a classification model with a help of a machine learning algorithm. Jupyter notebooks is used as a software tool in this project.

A. Data Pre-manipulation :

This particular fragment of the report will graft the data, and examine the standard of data cand then trim and clean the given dataset for analysis as mentioned in figure 3.1 .

B. Data Graphing :

The data set collected for predicting given data is split into the training set and a testing set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using machine learning algorithms is applied to the Training set and based on the test result accuracy. Here matplotlib is used for data visualization.

C. Designing the classification model:

This classification model is made up of four algorithms of Machine Learning such as **Logistic Regression, Random Forest, Decision Tree Classifier, and Naïve Bayes.**

i.   **Logistic Regression algorithm**

It is a classification algorithm in the technique of machine learning which is used to detect the possibilities of a categorical dependent variable. The purpose of logistic regression is to find the exact model to elaborate the association between the bifurcate characteristic of interest (measured variable = reply or final output variable) and a set of variables. The final output of confusion matrix is mentioned in the figure C i .
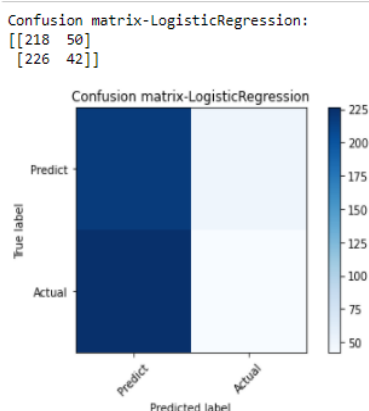
$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

figure C i

ii.   **Random Forest algorithm:**

It is a type of supervised machine learning algorithm which is completely about the ensemble learning. It is a type of learning where you connect various types of algorithms or similar algorithm for more number of times to create a more effective dectecting model .This algorithm gathers many more algorithms of the similar type i.e. multiple decision *trees,* resulting in a *forest of trees,* hence the name

"Random Forest". The random forest algorithm can be used for both regression and categorial chores. The output of confusion matrix is mentioned in the figure C ii .

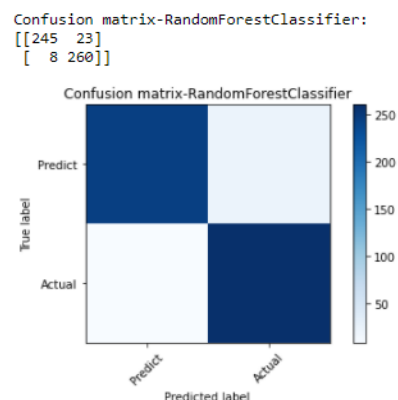$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

figure C ii

iii.   **Decision Tree Classifier algorithm**

Decision-tree algorithm is one of the most effective category and well familiar category of supervised learning algorithms. It is suitable for continuous and categorial variables of the output . Decision tree constructs classification or regression models and assembles it in a formation of a tree. It breaks down the sets of data into inferior and inferior subsets while concurrently an co-related decision tree is incrementally designed. A root node has two or more branches and a leaf node which represents a classification or a categorial decision. It is designed in a top-down recursive approach and in a divide-and-conquer manner. Each and every attributes included in this tree should

be categorical. The output of confusion matrix is mentioned in the figure C iii .
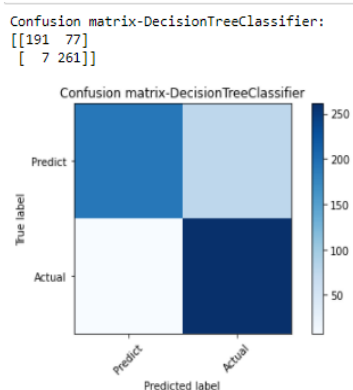
```
Confusion matrix-DecisionTreeClassifier:
[[191  77]
 [  7 261]]
```


figure C iii

iv.   Naive Bayes algorithm:

The Naive Bayes algorithm is an intuitive method that uses the probabilities of each attribute belonging to each class to make a prediction. Naive bayes simplifies the calculation of probabilities by assuming that the probability of each attribute belonging to a given class value is independent of all other attributes. This is a strong assumption but results in a fast and effective method. The probability of a class value given a value of an attribute is called the conditional probability. By multiplying the conditional probabilities together for each attribute for a given class value, we have a probability of a data instance belonging to that class. To make a prediction we can calculate probabilities of the instance belonging to each class and select the class value with the highest probability. Naive Bayes classifier assumes that the effect of a particular feature in a class

is independent of other features. For example, a loan applicant is desirable or not depending on his/her income, previous loan and transaction history, age, and location. The output of confusion matrix is mentioned in the figure C iv .
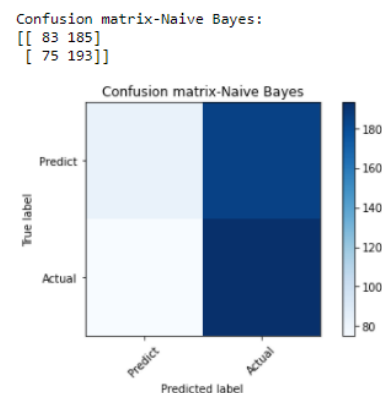
```
Confusion matrix-Naive Bayes:
[[ 83 185]
 [ 75 193]]
```


figure C iv

D . Flask Deployment

Flask is a micro web framework written in Python. The accuracy of all the algorithm is compared as mentioned in figure 3.2 and the best algorithm (Random Forest) is deployed using flask into a webpage . It is classified as a micro-framework because it does not require particular tools or libraries.

5. TESTING

5.1 SOFTWARE TESTING :

We've performed software testing to discover an unprecedented error which could have damaged the future of the software . This gives the record of the process of test planning with significant input from peers.

5.2 A/B TESTING :

We've performed this testing to test and improve machine learning models. We used this technique to decide whether a new model has an

improvement over a current model. We done it to checkout the improvement in its accuracy .

5.3 STRESS TESTING :

Once we deployed as a webpage , we had nearly 150 logins concurrently . Hence this resulted in a user-friendly webpage .

6. RESULTS AND DISCUSSION

Hence the data sets are compared with various algorithms and the level of stress is detected accurately at an early stage where it prevents the future cause of stress. During testing we received suggestions from various users hence improvised by including precaution measures for the people in stress.

7. CONCLUSION AND FUTURE WORK :

The analytical process started with data purifying, lost values , probing interpretation, and finally the model is being built and evaluated . The best efficiency is on the public test set is a higher accuracy score will be found. This application can help to find the Prediction of mental health and future work for this project can be

- Connecting the mental health prediction with the cloud model.
- Optimizing the work to implement in an Artificial Intelligence environment

8. REFERENCES

[1] Jung Yuchae, Yong Ik. Multimedia Tools and Applications, 76 (9) (2020), pp. 11305-11317 View PDF CrossRefView Record in ScopusGoogle Scholar

[2] Norizam, Sulaiman. Determination and classification of human stress index using the nonparametric analysis of EEG signals. Diss. UniversitiTeknologi MARA, 2020. Google Scholar

[3]http://www.mindgarden.com/documents/Perceived StressScale.pdfGoogle Scholar

[4] Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms (2020). https://doi.org/10.1016/j.procs.2020.03.442

[5] Stress Detection Using Machine Learning Algorithms(2020) . https://www.journals.resaim.com/ijresm/article/download/171/154/315

[6] Mental Stress Level Prediction and Classification based on Machine Learning (2020) https://ieeexplore.ieee.org/document/9588803