



e-ISSN:2582-7219



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 6, Issue 6, June 2023



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.54



6381 907 438



6381 907 438



ijmrset@gmail.com



www.ijmrset.com



Mental Health Prediction for Human

Dr. T. GEETHA, U. SRIAISHWARYA, M. THARANI

Head, Department of MCA, Gnanamani College of Technology, Namakkal, India

Assistant Professor, Department of MCA, Gnanamani College of Technology, Namakkal, India

Student, Department of MCA, Gnanamani College of Technology, Namakkal, India

ABSTRACT - Mental illness is a health problem that undoubtedly impacts emotions, reasoning, and social interaction of a person. These issues have shown that mental illness gives serious consequences across societies and demands new strategies for prevention and intervention. To accomplish these strategies, early detection of mental health is an essential procedure. The widely used algorithms in the field of machine learning such as random forest, logistic regression, Knn algorithm have been utilized to forecast and categorize the future events.

KEYWORDS - Mental Health Prediction, Machine Learning Algorithm, Depression, Anxiety, Social Well-being, Mental health, Stress.

I. INTRODUCTION

ML is becoming more capable for disease diagnosis and also provides a platform for doctors to analyze a large number of patient data and create personalized treatment according to the patient's medical situation. In this article, we are going to predict the mental health for unemployment and mental illness survey of using machine learning techniques. Mental illness is a health problem that undoubtedly impacts emotions, reasoning, and social interaction of a person. These issues have shown that mental illness gives serious consequences across societies and demands new strategies for prevention and intervention. To accomplish these strategies, early detection of mental health is an essential procedure. Machine learning is a technique that aims to construct systems that can improve through experience by using advanced statistical and probabilistic techniques. It is believed to be a significantly useful tool to help in predicting mental health. It is allowing many researchers to acquire important information from the data, provide personalized experiences, and develop automated intelligent systems. The widely used algorithms in the field of machine learning such as random forest, linear regression, Knn algorithm have been utilized to forecast and categorize the future

II. DATA ANALYSIS AND VISUALIZATION

2.1 Data Cleaning:

The data can have many irrelevant and missing parts. To handle this part, data cleaning is done. It involves handling of missing data, noisy data etc.

- **Missing Data:** This situation arises when some data is missing in the data. It can be handled in various ways.
- **Ignore the tuples:** This approach is suitable only when the dataset we have is quite large and multiple values are missing with in a tuple.
- **Fill the Missing values:** There are various ways to do this task. You can choose to fill the missing values manually, by attribute mean or the most probable value.

2.2 Data Analyzation:

Data analysis involves manipulating, transforming, and visualizing data in order to infer meaningful insights from the results. Individuals, businesses, and even governments often take direction based on these insights. Data analysts might predict customer behavior, stock prices, or insurance claims by using basic linear regression. They might create homogeneous clusters using classification and regression trees (CART), or they might gain some impact insight by using graphs to visualize a financial technology company's portfolio. Until the final decades of the 20th century, human analysts were irreplaceable when it came to finding patterns in data.



2.3 Data visualization:

Data visualization is the representation of data through use of common graphics, such as charts, plots, info graphics, and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand. Data visualization can be utilized for a variety of purposes, and it's important to note that is not only reserved for use by data teams. Management also leverages it to convey organizational structure and hierarchy while data analysts and data scientists use it to discover and explain patterns and trends. Data visualization is a critical step in the data science process, helping teams and individuals convey data more effectively to colleagues and decision makers. Teams that manage reporting systems typically leverage defined templats views to monitor performance.

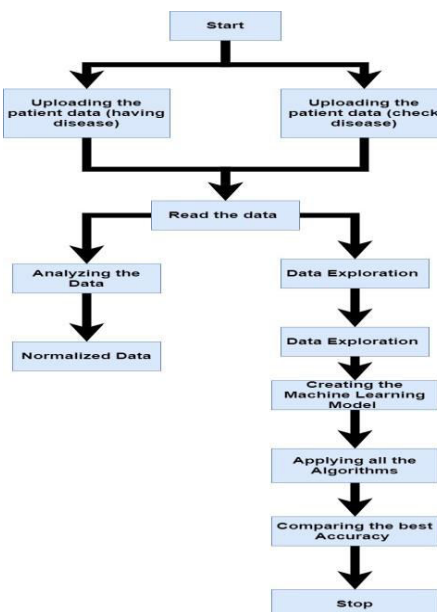
III. METHODOLOGY

3.1 THE DATASET

3.2 THE MACHINE LEARNING MODEL

The machine learning model is trained by iteratively modifying the strengths of the connections so that given inputs map to the correct response.

IV. PROPOSED SYSTEM





4.1 RANDOM FOREST:

Random forest is the most famous and it’s considered as the best algorithm for machine learning.it is a supervised learning algorithm. To achieve more accurate and consistent prediction, random forest creates several decision trees and combines them together. The major benefit of using its ability to solve both regression and classification issues.

4.2 LINEAR REGRESSION:

Linear Regression is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.

4.3 K-Nearest Neighbors (KNN):

K-Nearest Neighbors algorithm is a simple, supervised machine learning algorithm that can be used to solve problems. Here we are using for predicting Mental health prediction using a dataset acquired from Kaggle. Moreover, it has a major drawback of becoming significantly slows as the size of that data in use grows.

V. EVALUATION MATRICES

5.1 CONFUSION MATRIX:

A confusion matrix, also known as an error matrix, is a table that is often used to describe the performance of a classification model (or “classifier”) on a set of test data for which the true value share known. It allows the visualization of the performance of an algorithm. It allows easy identification of confusion between classes e.g., one class is commonly mislabeled as the other.

5.2 COVARIENCE MATRIX:

The covariance matrix is used to capture all the possibilities of how the variables present in a given dataset can change together. The function used to calculate the covariance matrix in python is called covariance function denoted by **cov** ().

5.3 CORRELATION MATRIX:

A correlation matrix is a table that contains correlation coefficients for several variables. The correlation between two variables is represented by each cell in the table. The value ranges from **-1 to 1**.

5.4 AUC_ROC CURVE:

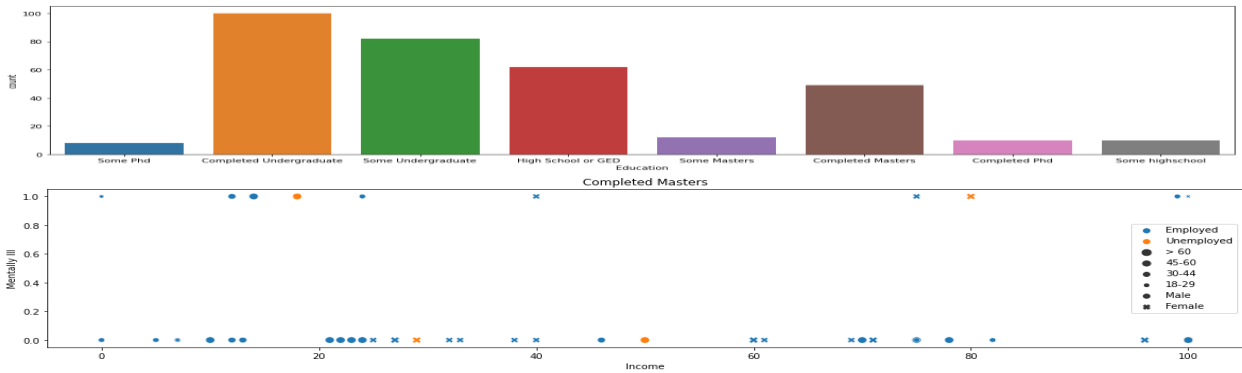
The curve is plotted between two parameters, which are: True Positive Rate or TPR False Positive Rate or FPR AUC is known for Area Under the ROC curve.

VI. OUTPUT

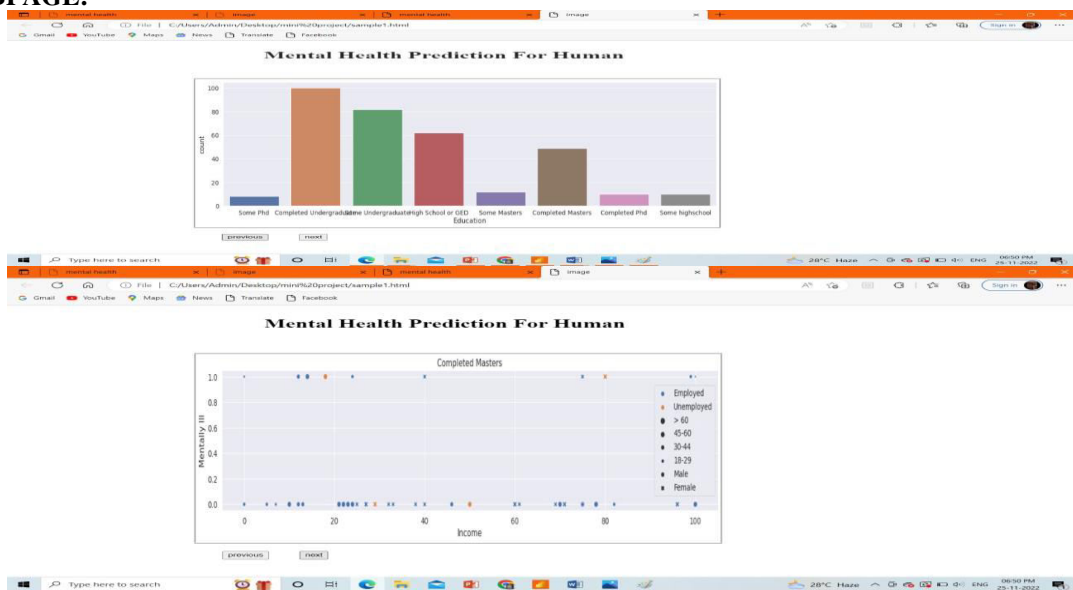
6.1 PLOT VISUALIZATION:

	I am currently employed at least part-time	MentallyIll	Education	I have my own computer separate from a smart phone	I have been hospitalized before for my mental illness	How many days were you hospitalized for your mental illness	I am legally disabled	I have my regular access to the internet	I live with my parents	I have a gap in my resume	...	Obsessive thinking	Mood swings	Panic attacks	Compulsive behavior	Tiredness	Age	Gender
1	1	1	Some Phd	1	0	0.0	0	1	0	0	...	0.0	0.0	1.0	0.0	1.0	18-29	Male
2	1	0	Completed Undergraduate	1	0	0.0	0	1	0	0	...	0.0	0.0	0.0	0.0	0.0	30-44	Male
3	0	0	Some Undergraduate	1	0	0.0	0	1	1	1	...	0.0	0.0	0.0	0.0	0.0	30-44	Male
4	1	1	Completed Undergraduate	1	1	35.0	1	1	0	1	...	1.0	1.0	1.0	1.0	1.0	30-44	Male
5	1	0	High School or GED	1	0	0.0	0	1	1	0	...	0.0	0.0	0.0	0.0	0.0	30-44	Male

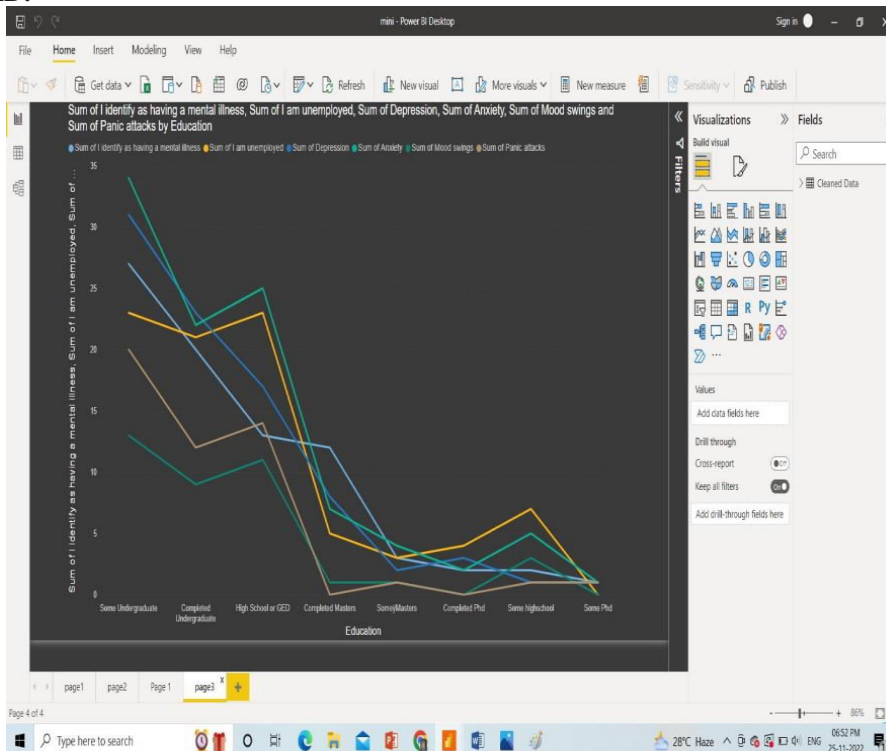
5 rows x 31 columns



6.2 WEBPAGE:



6.3 DASHBOARD:





VII. CONCLUSION

Without obvious indicators for mental illness such as hospital days and clear conditions, random forest with max_depth of 5 achieved 82% accuracy with a ROC-AUC of 0.9899. The clearest indicators of mental illness are income, education, gaps in resume, age, tiredness, lack of concentration. With clear indicators such as depression, anxiety, etc., random forest with max_depth of 10 achieved 94% accuracy and ROC-AUC of 0.997. Depression, anxiety, hospitalized days, and education were the most important features. As a predictor for unemployment, mental health considerations did play a significant role. Removing features such as current employment status and income, random forest with max_depth of 15 achieved 82% accuracy and ROC-AUC of 0.98.

REFERENCES

1. <https://www.javatpoint.com/machine-learning-algorithms>.
2. <https://www.kaggle.com/datasets/michaelacorley/unemployment-andmental-illness-survey>.
<https://github.com/cdodiya/Mental-Health-Prediction-using-Machine>. <https://www.javatpoint.com/data-preprocessing-machine-learning>.
3. <https://www.javatpoint.com/machine-learning-random-forest-algorithm>. <https://www.javatpoint.com/linear-regression-in-machine-learning>.
4. <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machinelearning>.
https://www.tutorialspoint.com/javascript_web_development/index.asp.
5. <https://learn.microsoft.com/en-us/power-bi/create-reports/servicedashboards>.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor
7.54

ISSN

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com