Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024

MENTAL HEALTH PREDICTION USING MACHINE LEARNING

Dr. S. N. TirumalaRao M.Tech,Ph.D., Dept. of Computer Science & Engineering Narasaraopet Engineering College Narasaraopet,India nagatirumalarao@gmail.com
Chetan Kumar Adugula Computer Science & Engineering Narasaraopet Engineering College Narasaraopet, India Chetanxe@gmail.com
Sharan Kumar Dugge Computer Science & Engineering Narasaraopet Engineering College Narasaraopet, India duggesharankumar@gmail.com
Naga Manikanta Thota Computer Science & Engineering Narasaraopet Engineering College Narasaraopet, India thotanagamanikanta01@gmail.com

Abstract-
Mental illness is becoming more widely acknowledged as a significant health issue these days. Patients with mental health issues are becoming more prevalent. As a result, we must devise a strategy to address this issue. Creating machine learning models that can forecast users' mental illnesses—that is, those with poor mental health—is the primary objective of this research effort. Here, by using some of the machine learning models, we were able to provide some insight into this issue. We used datasets containing the individual mental health information to create our models. Pre-processing was also necessary because the dataset can include noisy or erroneous data. On the data, several machine learning methods are used.

Keywords-
Mental Health, Self-Analysis, Machine Learning, Logistic Regression, Decision Tree, Radom Forest, Support Vector Machine, XG Boost Classifier, ADA boost, Gradient boost

I. INTRODUCTION

Mental health is one of the leading causes to suicide. As most of the victims know a little about their mental health, it goes unseen. Every year more than 800,000 individuals die due to Depression, or some mental health disorder and this number is not a Mostly youth and have a lot of potential to become something and give society a lot [1].

To achieve success in such cases, the potential cases must be sorted before even they know. Mental health has a lot to do with the amount earned and the way it has been spent [2]. Our spending and liabilities tell a lot about our potential mental health.

Our spending and liabilities tell a lot about our potential mental health[3]. examining previous cases, in which it has considered how much individual do earn and there spending habits or where there most chunk of money goes, if they use tobacco, alcohol, medical expenses, Social expenses, Education expenses, number[4] of time they eat meat or fish, slept hungry or not, enough food for tomorrow, Doctor consulting and other Also the one or which could be ignored [5]. Most of the individuals are between the age between 15-24. Which is mostly youth and have a lot of potential to become something and give society a lot [6].

To achieve success in such cases, the potential cases must be sorted before even they know. Mental health has a lot to do with the amount earned and the way it has been spent [7]. Our spending and liabilities tell a lot about our potential mental health. examining previous cases, in which it has considered how much individual do earn and there spending habits or where there most chunk of money goes, if they use tobacco, alcohol, medical expenses, social expenses, Education expenses, number of times they eat meat or fish, slept hungry or not, enough food for tomorrow, Doctor consulting and other kind of liabilities [8].

II.LITERATURE SURVEY

This study evaluated the practicality of many machine learning algorithms that classify the information into distinct mental health categories. This framework was developed to assess an individual's mental

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024

health status, and models for evaluation were constructed using this framework in mind.

Bhattacharyya et al. [10] tells that A person's enthusiastic, mental, and social well-being are all indicated by their mental health. It determines how a person thinks, feels, and responds to situations. A person's ability to work profitably and reach their maximum potential depends on their mental health. These remarks were made in 2018.

Shatter et al. [11] A research investigation of machine learning and its potential connection to mental health difficulties has been prompted by the rise in mental health concerns and the need for effective medical care. This research offers a recent accurate assessment of machine learning techniques for predicting mental health problems.

Jetli chug et al [12] invented the Behavioral health disorders, specifically distress, are the kinds of health issues that many people are unaware of. One cannot possibly receive treatment for something they are unaware of. Therefore, identifying a person who may have a wellness problem is the first step in avoiding them.

III.METHODOLOGY

A. Data Collection and Preprocessing:

Getting the data is a crucial step in any machine learning project, as the quality of the data affects how well the model works. In this study, we obtained our data from Kaggle, a popular website where data scientists share datasets. After collecting the data, which comprised over 2000 records and 21 attributes information, we uploaded it to Google Colab, an online platform for analyzing data and performing machine learning tasks. We utilized four datasets from Kaggle, each containing a different number of records

Pre-processing of Dataset:

Before we could use the data, we had to clean it up by getting rid of any missing information and unusual values. This makes the data ready for training and testing our models. Here we are going to show the missing values if the through the heat map.

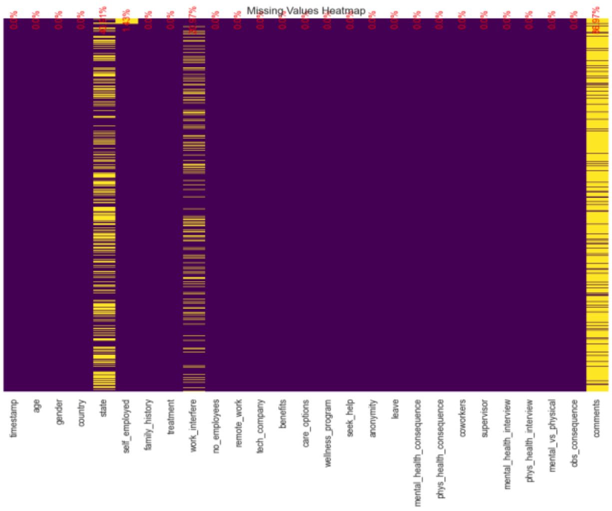

Fig.2. Heatmap

Through this heat map we are going to showing the missing date in the data set after pre-processing steps. The missing data attribute will be shows the heat bars in the figure.

df.info (): Which is used to see the values are null or not null.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024

0	timestamp	1259 non-null	object
1	age	1259 non-null	int64
2	gender	1259 non-null	object
3	country	1259 non-null	object
4	state	744 non-null	object
5	self_employed	1241 non-null	object
6	family_history	1259 non-null	object
7	treatment	1259 non-null	object
8	work_interfere	995 non-null	object
9	no_employees	1259 non-null	object
10	remote_work	1259 non-null	object
11	tech_company	1259 non-null	object
12	benefits	1259 non-null	object
13	care_options	1259 non-null	object
14	wellness_program	1259 non-null	object
15	seek_help	1259 non-null	object
16	anonymity	1259 non-null	object
17	leave	1259 non-null	object
18	mental_health_consequence	1259 non-null	object
19	phys_health_consequence	1259 non-null	object

Fig.3. Dataset info

After the Categorical string values will be transform into the numerical values the correlation of the each attribute will be represent in the correlation heat map.

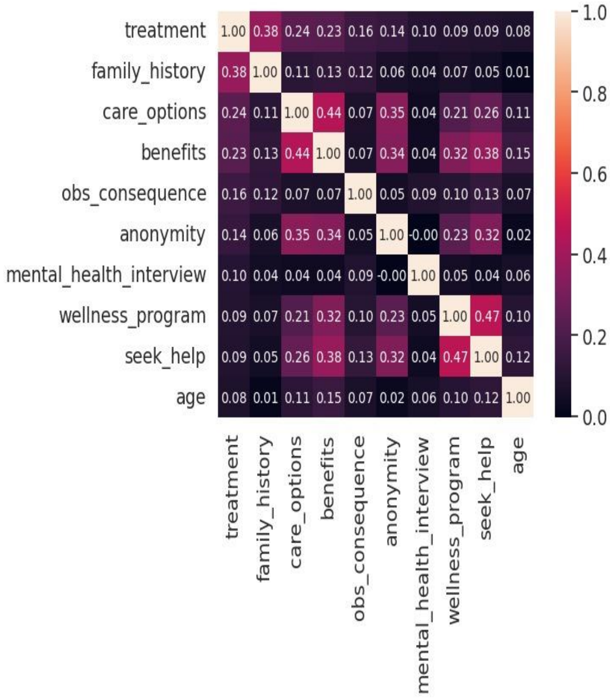

Fig.4. Correlation Heatmap

df. isnull (): The method which is used to find out the missing values of the attributes present in data set.

```
timestamp                      0.00
age                            0.00
gender                         0.00
country                        0.00
state                         40.91
self_employed                  1.43
family_history                 0.00
treatment                      0.00
work_interfere                20.97
no_employees                   0.00
remote_work                    0.00
tech_company                   0.00
benefits                       0.00
care_options                   0.00
wellness_program               0.00
seek_help                      0.00
anonymity                      0.00
leave                          0.00
mental_health_consequence      0.00
phys_health_consequence        0.00
coworkers                      0.00
supervisor                     0.00
mental_health_interview        0.00
phys_health_interview          0.00
mental_vs_physical             0.00
obs_consequence                0.00
comments                      86.97
```

. Fig.5. checking for missing values.

B. Data Visualization

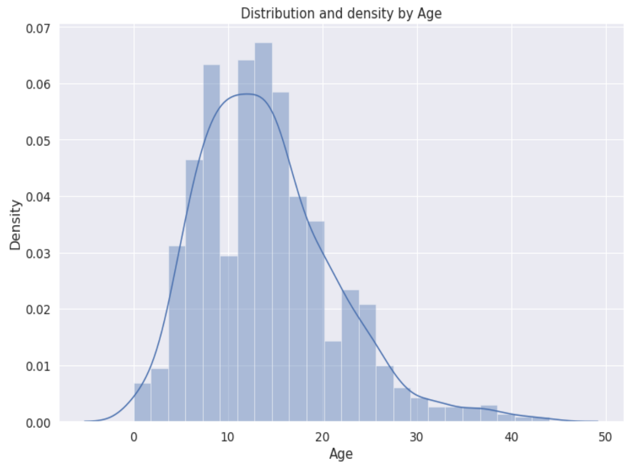

Fig 7. Density graph

The Age column in relation to density is displayed in the plot above. In our dataset, we can see that the density increases between the ages of 10 and 20.

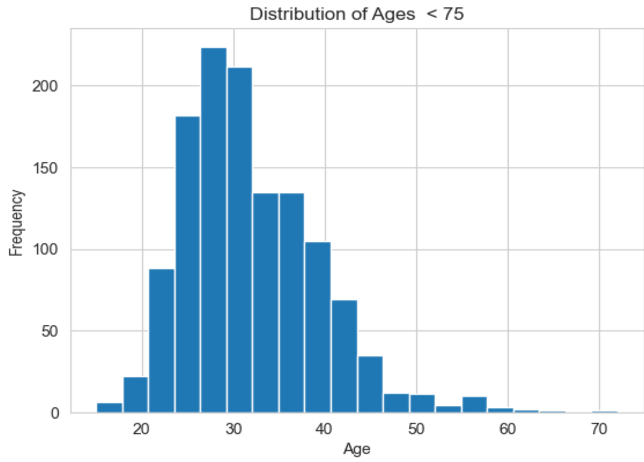

Fig .8. Bar graph

In Bar graphs for data visualization are essential to the "mental health prediction using Machine Learning Algorithms" project because they provide various aspects of the dataset and model performance.

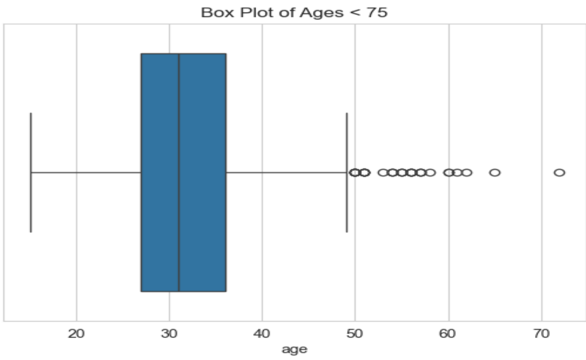

Fig .9. Box plot

Box plots are a useful tool for displaying the collection of numerical data values in between the different age groups, particularly when compare them across several groups.

C. Model Assessment and Selection:

This phase involves selecting a suitable machine learning algorithm and training it on the preprocessed data. By fine-tuning its parameters to minimize the discrepancy between the expected and actual outputs in the training set, the model is trained. The model is tested on a different approval dataset to gauge its performance after training. The selection of evaluation features is contingent upon the type of conflicts and the performance standards. Accuracy, precision, cross-validation, and F1score are examples of common evaluation measures. By changing its settings or using various algorithms, the model can be further refined in light of the evaluation results. To improve the model's performance on fresh, untested data, this step is crucial. The model can be used to make predictions or choices in a production setting after it has been trained and assessed.

IV.RESULTS AND CONCLUSION

Deployment

The selected model will be dumped into the app.py And takes the input from the user as a questioner and predict output.

Fig.10. output screen

Accuracy comparison between the different algorithms.

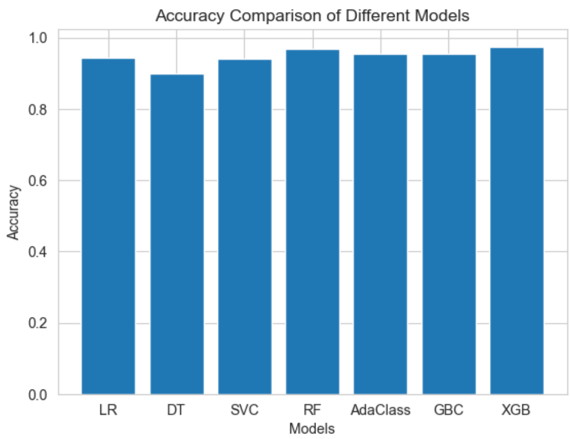

Fig.12. Accuracy comparison graph

The Above diagram represents the accuracy comparisons graph.

And by comparing the all the accuracies bars the graph contains the highest contribution of XGboost than any others.

TABLE -1 Accuracy of different algorithms

Algorithms	Accuracy
Random Forest	0.97
Decision Tree	0.96
SVC	0.90
Logistic Regression	0.94
ADABoost	0.95
Gradient Boost	0.95
XGboost	0.1

The performance of various algorithms was evaluated based on their accuracy scores for a specific task. Random Forest emerged as the top performer with an accuracy of 0.97, demonstrating its effectiveness. Following closely behind was the Decision Tree algorithm with an accuracy of 0.96, indicating its suitability for the task. In contrast, the models including ADBoost, Gradient boost, displayed similar accuracy scores of 0.95, which, although lower than the tree-based models, can still be valuable in certain contexts. XGBoost achieved an accuracy of 0.1, respectable but falling behind the ensemble methods. These results emphasize the effectiveness of ensemble methods for the task, particularly Random Forest and Decision Tree, while also highlighting the importance of considering but after comparing the performance of XGboost we are going to consider that. To sum up, this study investigated the use of popular machine learning algorithms, such as ADABoost, GRADIENT Boost, SVM, Random Forest, Decision Tree, Logistic Regression, and XGBoost, to calculate the likelihood that a child will experience mental illness. The acquired findings show that the algorithms are capable of producing accurate predictions, with XGBoost and Random Forest surpassing all other models. These two models function in terms of productivity and accuracy. The factors taken into account in this predictive model include age, self-employment, work-life balance, family history, coworkers, and the effects on one's physical and mental health. All things considered, this work demonstrates the promise of machine learning algorithms and how they might improve forecast accuracy and dependability. By adding more intricate characteristics, investigating novel and alternate techniques, and assessing the model's performance on sizable datasets, future research can build on this work.

REFERENCES

[1] Stevie Chancellor, Eric PS Baumer, and Munmun De Choudhury. Who is the human in human-centered machine learning: The case of “predicting mental health from social media”. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW):1–32, 2019.

[2] Sarah Graham, Colin Depp, Ellen E Lee, Camille Nebeker, Xin Tu, Ho-Cheol Kim, and Dilip V Jeste. "Artificial intelligence for mental health and mental illnesses" an overview. Current psychiatry reports, 21(11):1–18, 2019
[3] T Nagar. "Prediction of mental health problems among children using machine learning techniques"
[4] K. Sowmya Fathima, S. Barker and S. Kulkarni, "Analysis of mental illness prediction using data mining technique", International Research Journal of Engineering and Technology (IRJET), vol. 07, no. 05, May 2020.
[5] U. Srinivasulu Reddy, Aditya Vivek Thota and A. Dharun, "Machine learning techniques for stress prediction in working employees", *2018* IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018.
[6] Suresh, A Stephen M. Schueller, Adrian Aguilera and David C. Mohr, "Ecological momentary interventions for depression and anxiety", Depression and Anxiety, vol. 34, no. 6, pp. 540-545, 2017
[7] U. Srinivasulu Reddy, Aditya Vivek Thota and A.Dharun,"Machine learning techniques for stress prediction in working employees", 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018.
[8] Suresh, G., A. Senthil Kumar, S.Lekashri, and R.Manikandan. "Efficient mental health System Using Machine Learning." International Journal of Mental health 10, no. 1 (2021): 906-914
[9] D. Graham, Sarah, et al. "Artificial intelligence for mental health and mental illnesses" an overview." Current psychiatry reports 21.11 (2019): 1-18
[10] Bhattacharyya, R., & Basu, S. (2018). India Inc looks to deal with "rising stress in employees". Retrieved from 'The Economic Times'
[11] Shatter, Adrian BR, Delyse M. Hutchinson, and Samantha J. Teague. "Machine learning in mental health" a scoping review of methods and applications. Psychological medicine (2019)
[12] M. Hamilton, "Development of a rating scale for primary depressive illness," *British Journal of Social and Clinical Psychology*, vol. 6, no. 4, pp. 278–296, 1967.
[13] G. Cho, J. Yim, Y. Choi, J. Ko, and S.-H. Lee, "Review of machine learning algorithms for diagnosing mental illness," *Psychiatry Investigation*, vol. 16, no. 4, pp. 262–269, 2019.
.[14] M. Hamilton, "Development of a rating scale for primary depressive illness," *British Journal of Social and Clinical Psychology*, vol. 6, no. 4, pp. 278–296, 1967.