Mental Health Predictor : Using Machine Learning Algorithms

Ms. Mudit Ranjan Shrivastava
B.Tech (Scholar), Department of Computer Science & Engineering, Shri Shankaracharya nstitute of Professional Management and Technology, Raipur, Chhattisgarh, India
Email: mudit.shrivastava@ssipmt.com

Mr. Devbrat Sahu
Assistant Professor, Department of Computer Science and Engineering Shri Shankaracharya Institute Of Professional Management and Technology, Raipur, Chhattisgarh, India
Email: devbratsahu89@gmail.com

Ms. Varun Mishra
B.Tech (Scholar), Department of Computer Science & Engineering, Shri Shankaracharya Institute of Professional Management and Technology, Raipur, Chhattisgarh, India
Email: varun.mishra@ssipmt.com

Ms. Piyush Agarwal
B.Tech (Scholar), Department of Computer Science & Engineering, Shri Shankaracharya Institute of Professional Management and Technology, Raipur, Chhattisgarh, India
Email: piyush.agarwal@ssipmt.com

Abstract*— *Early detection of mental health problems improves patients' quality of life and allows professionals to treat them more effectively. Being mentally, emotionally and socially well is important. It affects how you feel, think and act. Health, whether physical or mental, plays an important role in human life. The effectiveness of these machine learning techniques in identifying mental health-related issues was evaluated using . There are five machine learning algorithms: logistic regression, K-NN classifier, decision tree classifier, random forest, and stacking. After comparing and practicing these strategies, we found the stacking method to be the most accurate, with a prediction accuracy of 81.35%.

Keywords*—*Mental Health, current state, Machine Learning, data, depression

I. INTRODUCTION

A person's mental health depends on both their current state of mind and how they interact with the world around them. Insanity is caused by abnormalities in brain chemistry. It is important to assess different categories of mental health at different times to avoid serious illness. Depression and anxiety are serious disorders that affect the health of people around the world. Inferring a person's mental state based on their behavior and appearance is an advanced psychological science that has not yet been mechanized. Anxiety and depressive disorders have far-reaching effects on health and well-being. They are suffering from various contagious diseases such as HIV and tuberculosis. Smartphones, social media, neuroimaging, and wearable technology have enabled mental health medical professionals and researchers to acquire vast amounts of data. Machine learning has helped areas such as natural language processing, speech recognition, computer vision, and artificial intelligence. People lose access to economic and social opportunities, and their quality of life is adversely affected. Technological smartphones, social media, neuroimaging, and wearable technology have enabled medical professionals and mental health researchers to acquire large amounts of data quickly.

II. LITERATURE SURVEY

To classify information into different mental disorders, this study evaluated the effectiveness of eight different machine learning algorithms. According to their findings, Multiclass Classifier, Multilayer Perceptron, and LAD Tree are three classifiers that provide more accurate results than others. The paper articulates the analysis of mental health in terms understandable by a variety of audiences. They developed a framework for assessing a person's mental health and used it to create a predictive model. Clustering techniques were also used to count the number of clusters before building the model. The generated class labels were validated with MOS before being used to train the classifier. Testing has shown that KNN, SVM, and Random Forest perform very similarly. We also found that using the ensemble classifier significantly improved our ability to predict mental health with 90% accuracy. This research focuses on how machine learning can help better detect and diagnose psychiatric disorders such as schizophrenia, depression and Alzheimer's disease. Overall, machine learning has the potential to improve the effectiveness of clinical and research processes, while opening new perspectives on mental health and well-being. The main contribution of this work is the application of his ILIOU preprocessing method, which predicts various types of depression and significantly improves the performance of classification algorithms on similar datasets. Anticipating depression is essential so that patients receive the best treatment as soon as possible. In medicine, there is an increasing focus on artificial intelligence to aid in mental health research and applications. To realize the full potential of AI, a diverse community of mental health research and care professionals, including scientists, clinicians, patients and regulators, must connect and communicate. In this work, we used discourse analysis to evaluate data to better understand representational practices in human-centered

machine learning (HCML). From this, they discovered a collection of 55 interdisciplinary research findings that predict mental health status based on social media data. Their results demonstrate the importance of contrasting interaction discourses present across datasets for developing and endowing human agency. Their results show how five discourses lead to a paradoxical view of humans as objects and subjects, perhaps inadvertently marginalizing humans.

III. METHODOLOGY

Data collection, data cleaning, coding, covariance matrix detection, scaling and fitting, fine-tuning, model evaluation, accuracy determination, and data and outcome prediction are all part of the process of knowledge from the data process, as shown in Figure 1. Included in discovery. First, assume your dataset has 1259 entries and 27 columns. The next step is data cleansing. Here, inaccurate, incomplete, unnecessary or missing data is identified and updated, replaced or deleted as appropriate. As you can see, 3 columns contain missing data. The special value Not a Number or NaN is used to indicate cells with no value in data frames and numpy arraysEncoding the data is the next step. Use this categorical data coding approach when categorical features are marked as ordinal. Maintaining order in this situation is very important. In order to reflect the sequence, the encoding should. During label encoding, each label is converted to an integer value. Then find the covariance matrix. This is one of the key matrices in data science and machine learning. Contains details about joint movements (correlations) of features. The variances of the variables are on the main diagonal of the variance-covariance matrix, and the covariances are between each pair of variables at other positions in the matrix. The mean of each variable is included in the mean vector. Feature scaling puts the independent features of the data into a given range. Handle values, units, or magnitudes that vary significantly during data preprocessing. Next, we split the dataset into a training dataset and a test dataset. Next is the importance of functions. Feature selection is critical in machine learning, with fundamental techniques that direct the use of variables to those that are most practical and efficient for a particular machine learning system. Tuning is the next stage. Improving model performance while avoiding overfitting and excessive variation is the process of tuningIn machine learning, this is achieved by choosing appropriate hyperparameters. Hyperparameters can be compared to the "dial" or "buttons" of a machine learning model. Then evaluate the model using various machine learning techniques such as stacking, logistic regression, K-nearest neighbor classifier, decision tree classifier, and random forest classifier. Logistic Regression – A well-known machine learning algorithm that falls under the supervised learning method is logistic regression. By using an unbiased set of variables, this strategy can be used to predict a specific dependent variable. Use logistic regression to predict the output of a specific structured variable. Therefore, the results should be discrete or categorical. It can be 0 or 1, yes or no, true or false, etc., but instead of giving an exact value between 0 and 1, it gives a range of probability values between 0 and 1. 3.2. K-Nearest Neighbor Classifier A basic machine learning algorithm that uses supervised learning methods is called K-Nearest Neighbor. Existing cases and new cases/data can be compared with the K-NN approach. ANN is a nonparametric algorithm that makes no assumptions about the distribution or highlighted data. It also works with multiple classes. Decision tree classifier, version 3. The most common supervised machine learning technique used in data mining is decision trees. Decision trees provide a visual representation of statistical probabilities or sequences of events, actions, or outcomes.

3.4. Random Forest Classifier A method called random forest based on supervised machine learning can solve classification and regression problems. However, classification is its common use. It is known as a random forest because it aggregates a large number of decision trees into a "forest" and feeds it random features from the input data set.

3.5. Stacking Stacked Generalization This is an ensemble machine learning algorithm. Similar to bagging and boosting, it combines predictions from different machine learning models on the same dataset..

IV. RESULTS

In this study, five machine learning methods were found: random forests, decision trees and stacking, logistic regression, and k-nearest neighbor classifiers. We also assessed how well they were able to identify mental health issues. We first ran the classifiers on all 27 attributes extracted from the text documents, and then ran them on 8 additional attributes that were selected using a feature selection approach. The incidence of correctly identified test sets using a classifier is the accuracy of that classifier's specific test set. Figure 2 shows that stacking accuracy is higher than other classifiers. A classifier's ability to correctly classify the test dataset determines the accuracy of the classifier. We calculated this using the area under the receiver operating curve. A perfect test has an area of 1 in the ROC range, and a useless test has an area of 0.5. A plot of the five classifiers against ROC area score is shown in Figure 3-7. We found that the classifier outperformed the others in predicting mental health status, as her ROC range for all classifiers used was between 0.8 and 0.9.

V. CONCLUSION

There are many approaches to machine learning, so it's important to compare them all and choose the one that best fits your target domain. Today, in medicine, there are many specialized programs that can predict diseases in advance with considerable accuracy, allowing for effective and rapid treatment. The proposed study compared five different machine learning approaches used to classify datasets of various mental health problems. The results revealed that all five machine learning methods provided more accurate results. Accuracy for all classifiers exceeds 79%. Only relatively small datasets were used in this study. In the future, larger datasets can be used to apply research and improve accuracy.

REFERENCES

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

[2] Colin Depp, Xin Tu, Ho-Cheol Kim, Sarah Graham, Ellen E. Lee, Camille Nebeker, and Dilip V. Jeste. An overview of artificial intelligence for mental health and mental diseases. Psychiatry today, 21(11), 1–18, 2019.

[3] George Anastassopoulos, Christina Lymperopoulou, Mandani Ntekouli, Konstantinos Assimakopoulos, Dimitrios Galiatsatos, and Theodoros Iliou Method for predicting depression using machine learning by Iliou. 2019. Evolving Systems, 10(1):29–39 T Nagar

[4]. employing machine learning approaches to predict children's mental health issues.

[5] Samantha J Teague, Delyse M Hutchinson, and Adrian BR Shatte. A survey of methodologies and applications for machine learning in mental health. Mental health, 49(9):1426–1448, 2019.

[6] S Mohanavalli, N Bhalaji, and M Srividya. Using machine learning algorithms, behavioural modelling for mental health. Medical Systems Journal, 42(5), 1–12, 2018.