Fuzzy Systems and Soft Computing
ISSN : 1819-4362

AN ENSEMBLE DATA MINING TECHNIQUE FOR CLASSIFICATION AND PREDICTION OF MENTAL ILLNESS USING "MIND PREDICT" TOOL

Mrs. P. Divya Bharathi Research Scholar PG & Research Department of Computer Science Government Arts College, Nandanam, Chennai. pdbdivya19695@gmail.com
Dr. R. Thirumalai Selvi Associate Professor PG & Research Department of Computer Science Government Arts College, Nandanam, Chennai. sarasselvi@gmail.com

ABSTRACT:
The mental well-being of a person is their mental state. Chemical abnormalities in the brain are the cause of mental health problems. In order to be prepared for problems related to health, it is critical to keep an eye on the mental health of various populations. College students and working professionals make up the community. It's common knowledge that people of all ages and backgrounds are impacted by stress and bereavement. Serious mental health conditions like schizophrenia, bipolar disorder, and anxiety don't always show up right away; they frequently develop over time and have early-stage symptoms. If abnormal mental states are identified early in the disease, more care and therapy can be given, and such mental problems may be more successfully prevented. To increase their accuracy in recognizing mental health disorders, we examined the performance of four data mining techniques and a novel ensemble technique in this work. The K-NN Classifier, Random Forest, Decision Tree Classifier, and Logistic Regression are the data mining approaches. Other academics and practitioners looking to diagnose mental health concerns accurately using improved data mining algorithms that meet many accuracy requirements will benefit from this study. After that, we have put the most accurate algorithm into practice and later run the data through it. After being taught, this algorithm will be a model and will be designed and tested. A web app has been developed named 'Mind Predict' where visitors can complete the forms and receive the appropriate outcome.

Keywords: Mental Illness, Prediction, KNN Classifier, Random Forest, Decision Tree Classifier, Logistic Regression.

I.INTRODUCTION

An individual's mental health is reflected in both their overall disposition and state of mind. Unbalances in brain chemistry lead to mental illness. The state of someone's mind reveals their emotional and social well-being. The target population is working class, College Students, and High School students, i.e..., above 18 years of age. Here we Review the existing data mining techniques to analyze and predict mental health problems. Based on internet data sources, this research offers a critical evaluation and analysis of mental health detection. Data are collected from online available datasets. Several attributes have been considered in our study. Data Preprocessing eliminates the unnecessary attributes from the dataset to create useful and meaningful information. The application of the data mining methodologies utilized for the study has been successful in making the data mining process valuable, meaningful, and effective. Google Research produces a product called Colaboratory, or simply "Colab." Anyone may write and run arbitrary Python code over the internet using Colab, which is particularly useful for data mining, teaching, machine learning, and analysis. Different approaches are required for patients when their mental illness worsens. Early diagnosis and detection are so crucial.

The focus point of this paper is to discuss the implementation of data mining techniques for predicting mental health of a person. In this study, we estimate and compare some of the most widely used classification algorithms that tackle this problem. Four data mining techniques namely Logistic Regression, K-NN Classifier, Decision Tree Classifier, Random Forest, and a unique Ensemble Technique, evaluating how well they can detect mental health problems. Technological innovations like smartphones, social media, neuroimaging, and wearables have made it possible for medical professionals and mental health researchers to quickly and efficiently collect vast amounts of data. As a dependable method of examining these data, data mining has grown in popularity. The use of

sophisticated statistical and probabilistic methods in data mining enables computers to learn from data on their own. This makes it possible to identify data patterns more quickly and accurately and to make more precise predictions using the data sources.

II. LITERATURE SURVEY

I have studied more than 10 papers as follows,

[1] Vidit Laijawala and Hardik Jatta (2020) have suggested that a person's mental health data mining technologies can reveal information about their emotional, psychological, and social well-being. This study investigates the use of several data mining approaches to forecast an individual's mental health conditions. The authors created the data model using Naïve Bayes, Random Forest, and Decision Tree Data mining algorithm. Prediction of mental illness are based on this algorithm and based on this the type of mental illness have been predicted.

[2] Nor Safika Mohd Shafiee and Sofianita Mutlib (2020) applied Supervised Learning to examine current machine learning methods for predicting and analyzing mental health issues among college students.

[3] Rohizah ABC, Rahman, Khairuddin Omar (2020) provides a comprehensive evaluation and examination of machine learning techniques and data sources for mental health detection in online social networks (OSN). Here the datas have undergone feature selection and the support vector machine algorithm have been applied to provide mental health detection in online social networks.

[4] Juan Li (2021) discussed the state of urban immigrant children's mental health at the moment, as determined by data mining algorithms and cloud computing utilizing the K-means clustering algorithm paradigm. This report presents the mental health status of children who moved from rural to urban areas.

[5] Muhammed Shabhaz, Shahzad Ali (2019) discusses the early detection and diagnosis of Alzheimer Disease (AD) using numerous Machine Learning Algorithms. Alzheimer Disease is a widespread neuro degenerative disease which causes cognitive impairment. Patients require different care as their AD worsens. Early diagnosis and detection are so crucial. KNN, Decision Tree, Rule Induction, Naïve Bayes Algorithm, and Generalized Linear Model (GLM) are the methodologies employed here.

[6] Chang Su and Zhenxing Zu (2020) conduct a review of the literature on the use of deep learning algorithms in research on mental health outcomes. This review also identifies a number of current obstacles to the clinical application of the DL algorithm for routine treatment, as well as encouraging future paths in this area.

[7]Susel Gongora Alonso et al (2018) focused on a summary of the literature's current research projects that discuss data mining methods and algorithms for the field of mental health. Here, the suggested course of action is to use the suggested approach to create a prediction model of cognitive impairment in patients and identify important risk factors for illness.

[8] In addition to other e-health domains, Subhan Tariq and Nadeem Akhtar (2019) suggest that social media and big data analytics have gained popularity as ways to predict mental illness in patients. These methodologies include data acquisition, data preprocessing, feature extraction, selection, and classification.

[9] According to J. Ruiz de Miras, A.J. Ibanez-Molina, M.F. Soriano, and S. Iglesias-Parro (2023), machine learning techniques can aid in the diagnosis of schizophrenia.

[10] Tianlin Zhang, Kailai Yang, Shaoxiong Ji, Sophia Ananiadou(2023) provides a comprehensive survey of approaches to identify mental disease through emotion fusion.

[11] Ojasvi Rajeev Sharma, Jyoti Agarwal, and Shaurya Bhatnagar (2023) The purpose of the study was to determine the level of anxiety and the consequences it has on Indian university students.

Table 1: Study of existing systems

S.no	TITLE OF THE PAPER	AUTHOR AND YEAR	ABSTRACT	METHODOLOGY	FUTURE WORK OR RESEARCH GAP
01.	[1] Classification Algorithm based Mental Health Prediction using Data Mining	Vidit Lajjawala, Hardik Jatta - 2020	▪ Target population is working class above 18 years of age.	▪ Decision Tree Algorithm ▪ Random Forest Algorithm	▪ We are able to develop a system that forecasts a certain mental disorder.
02	[2] Prediction of Mental Health Problems among higher education students using Machine Learning	Nor Safika, Mohd Shafiee Sofianita- 2020 Dec	▪ Examine the current state of machine learning to forecast and analyze mental health issues.	▪ Supervised Learning Techniques	--
03	[3] Application of Machine Learning Methods in Mental Health Detection – Systematic Review	Rohizah ABD Rahman, Khairuddin Omar – 2020	▪ This work offers a critical evaluation and analysis of the identification of mental health issues in online social networks (OSNs).	▪ Feature Selection ▪ Support Vector Machine	▪ For this study to improve the precision and accuracy of mental health problem diagnosis, a full adoption of revolutionary algorithms and computational linguistics was necessary.
04	[4] Analysis of Mental Health of Urban Migrant children Based on Cloud Computing and Data Mining Algorithm models	Juan Li – Sep 2021	▪ This article examines the current state of mental health among children of urban	▪ K- means Clustering Algorithm Model	--

			migrants.		
05	Data Mining Algorithm and Techniques in Mental Health – Systematic Review	Susel Gongon Alonso – 2018	▪ An analysis of the literature pertaining to data mining algorithms and methodologies in the field of mental health.	--	▪ Procedures from the patient database for schizophrenia compare procedures rather than assessing the accuracy and performance of the outcome. ▪ The creation of a patient cognitive impairment prediction model is another avenue for future research that has been suggested.
06	Mental Health Prediction Using Data Mining – A Systematic Review	Vidit Lajjawala, Hardik Jatta - 2020	▪ The data is collected from online available datasets ▪ Target population is working class	1. Decision Tree Algorithm 2. Random Forest Algorithm **TOOLS USED:** 1. WEKA	▪ This algorithm will be tested and the model will be designed ▪ It would then be deployed on a webpage where users can fill the forms and get the result.
07	[5] Classification of Alzheimer Disease(AD) using Machine Learning Techniques	Muhammed Shahbaz, Shahzad Ali – 2019	▪ Alzheimer Disease(AD) is a widespread Neuro	▪ KNN ▪ Decision Tree ▪ Rule Induction	▪ Increasing the number of instances for EMCI and SMC classes

			degenerative disease which cause cognitive impairment	▪ Naïve Bayes ▪ Generalized Linear Model (GLM)	could further increase the accuracy of AD phases classification.
08.	[8] A Novel Co-Training based Approach for the Classification of Mental Illness using Social Media Post	Subhan Tariq, Nadeem Akhtar – 2019	▪ In addition to other e-health domains, social media and big data analytics have gained popularity as tools for predicting patients' mental illnesses.	▪ Data Acquisition ▪ Data Preprocessing ▪ Feature Extraction/ Selection ▪ Classification	▪ Use the suggested method to categorize certain domain segments based on the interests of the scientific community.
09.	[6]Deep Learning in Mental Health outcome Research – A Scoping Review	Chang Su, Zhenxing Zu – 2020	▪ This study aims to review previous studies on the use of the DL algorithm in mental health outcome research.	---	▪ This review also identifies a number of current obstacles to the clinical application of the DL algorithm for routine care, as well as encouraging future directions for the discipline.
10	A Survey on Big-data driven digital phenotyping of Mental Health	Yunji Liang, Xiaolong Zheng, Daniel D.Zeng – April 2019	▪ Examining several unresolved problems and their related solutions to support digital phenotyping of mental health	--	▪ Thorough large-scale evaluation is required for digital phenotyping's clinical uses.

There are numerous varieties of systems in use at the moment. Most of them use different methods to predict mental disease.

An online survey included in some of the current systems determines whether or not the user has a mental illness. These surveys are condition-specific; there is a different one for stress, another for depression, and so on.

These surveys are all accessible online for completion by anybody.

A few systems make use of chat bots to ask users questions and then analyze their responses in order to predict mental illness. Certain systems also employ image processing to track users' facial expressions and analyze their responses to particular questions in order to help make a more accurate diagnosis of mental illness.

Most of these questionnaires focus on a person's appearance and demeanor, but they omit any questions about their place of employment. Because of this, not much study has been done on mental illness and the job. Most of these systems concentrate on the general features of mental disease. They make up

the measures that are most frequently used to evaluate the result. There aren't many processes in place to deal with mental illness in the workplace and among employees.

III. RESEARCH METHODOLOGY

In order to increase the accuracy of four data mining strategies in recognizing mental health disorders, we examined their accuracy in this study. The K-NN Classifier, Random Forest, Decision Tree Classifier, and Logistic Regression are the data mining approaches.

3.1 DATA MINING ALGORITHMS

3.1.1 Decision Tree Algorithm

- Among the techniques for supervised learning are decision tree algorithms. It can be used to solve problems with regression and classification. When traits are noted on the tree's internal node and each leaf node uses the tree representation to match to a class label, the issue is resolved. For selection, the entropy or Gini value is utilized.

3.1.2 Random Forest Algorithm

• The random forest approach, as its name implies, uses a large number of interconnected decision trees. Our model predicts the class with the most votes out of all the classes that each tree in the random forest forecasts.

3.1.3 Logistic Regression

- Logistic regression is a key data mining technique that is part of the supervised learning strategy. This method makes predictions about a dependent variable based on a set of independent factors. Certain structured variables' outputs can be predicted using the results of logistic regression.

3.1.4 KNN Classifier

- A basic data mining algorithm called K-Nearest Neighbor is based on instruction learning. New cases and old cases are similar in the K-NN approach. KNN is a non-parametric algorithm that doesn't assume anything about the distribution or data it highlights. It is also compatible with several classes.

Result and Analysis for Classification Algorithm Accuracy

The collection mostly includes information about people who are employed. For improved prediction, there are 25 questions collected from the Kaggle Online Datasets.

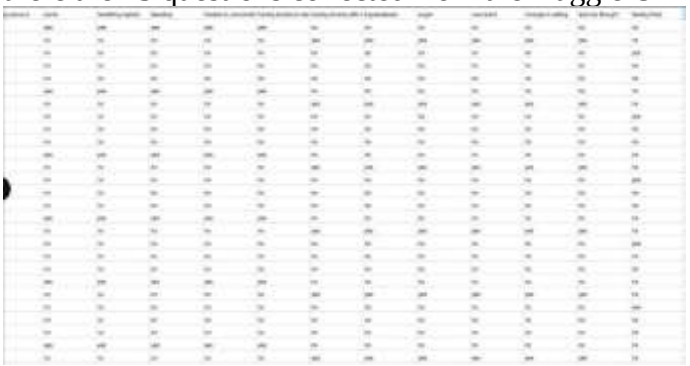

Figure 3: Dataset

The Dataset used is of the size of 5 X 25 (5 rows and 25 Columns).

To achieve our goals, we conducted a variety of trials.

The performance of the classification model is evaluated based on how well the classifier classified the data.

The confusion matrix makes use of the following terms:

(TP): Number of True Productive (records accurately identified as positive by the classifier).

(TUP): Number of True Unproductive (record properly identified as negative by classifier).

(FP): The quantity of false positives (records that the classifier mislabeled as positive).

(FUP): Amount of false negative records (classifier mistakenly tagged record as negative).

Accuracy, precision, recall, and specificity were the metrics we utilized to assess and contrast classifier performance.
Accuracy (as a percentage of all productive right answers):

$$\text{Accuracy} = \frac{TN+TP}{TN+FP+TP+FN}$$

Precision (proportion of correct productive observations):

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall (the percentage of positives that were accurately forecast to be productive)

$$\text{Recall} = \frac{TP}{TP+FP}$$

F1-Score

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision}+\text{Recall}}$$

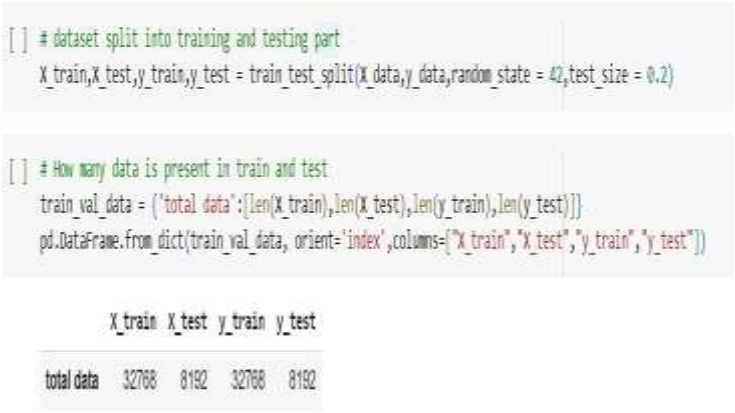

Figure 4: Test and Training Data

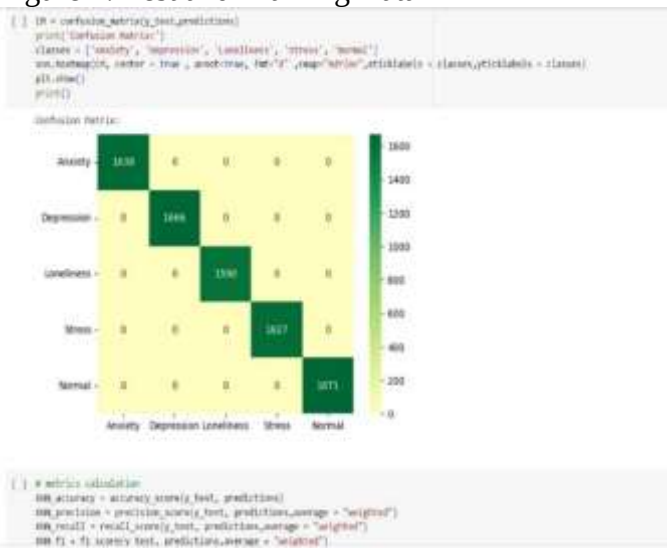

Figure 5: Confusion Metrics

Training using Decision Tree classifier algorithm

Figure 6: Training using Decision Tree Classifier Algorithm

Training Using RandomForestClassifier Algorithm

Figure 7: Training using Random Forest Classifier Algorithm

- ***We may develop a system in the future that can identify a particular mental illness a person has, but it will require a lot of data to be gathered***.

S.NO	ALGORITHM	ACCURACY (%)	PRECISION (%)	RECALL (%)	F1-SCORE (%)
01	K-NN Classifier	79	77	82	79
02	Logistic Regression	85	80	89	83
03	Decision Tree Algorithm	82	76	87	84
04	Random Forest	93	92	94	88
05	Ensemble Technique	96	94	97	90

Table 2: Classifiers and their Accuracy.

Table 2 presents the performance of different algorithms.

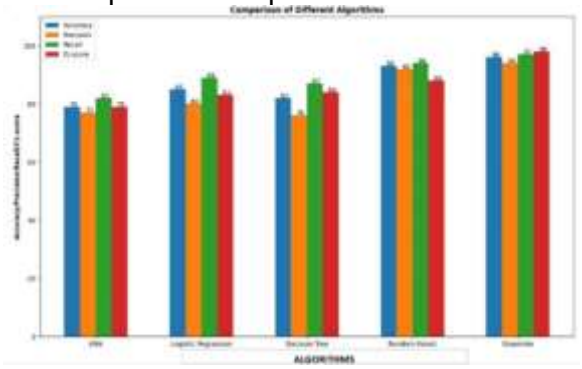

Figure 8: Comparison of Different Algorithm

Based on its high accuracy we have determined that Ensemble Technique is the most optimal algorithm.

We discovered that individuals who experience stress or depression at work ought to speak with a

mental health expert. On the other hand, an employee does not have mental illness if it has no effect on their ability to perform their job. Additionally, workers who occasionally experience problems at work or who have a family history of mental illness ought to get help.

IV. WEB APPLICATION TOOL ("MIND PREDICT") FOR PREDICTING MENTAL ILLNESS

The suggested approach in this case is to develop a web application tool to determine whether or not an individual is experiencing mental illness.

We have named the Web application tool as Mind Predict.

To determine who was experiencing mental illness, a series of questionnaires was utilized.

Not all of the parameters are helpful for the prediction because the dataset also includes information gathered for a questionnaire. As a result, the parameters we have chosen are appropriate and are displayed in the table.

QUESTIONS	ANSWERS →				
Feeling nervous	Yes	No	No	No	No
panic	Yes	No	No	No	No
Breathing rapidly	Yes	No	No	No	No
Sweating	Yes	No	No	No	No
Trouble in concentration	Yes	No	No	No	No
Having trouble in sleeping	Yes	No	No	No	No

Table 3: Sample questions provided as inputs

Our dataset has several factors to predict an employee's mental health, as Table 3 illustrates. The majority of data only have two attributes (Yes, No), The majority of the data is expressed as Yes or No values, indicating whether or not a person should seek treatment.

The questions and the datas were taken from online data source Kaggle for the implementation of our web app tool

A series of twenty-five questions were extracted from online resources and presented on the web application so that the user may enter their symptoms and receive a diagnosis of mental disease or not.

The coding were done in Python using GoogleColab.

This web app tool has a Login page where a user login with their username and password. Then there will be a web page displayed with a set of questions in which a user can answer and get the appropriate output.

Figure 1: Login page of Web Tool

Figure 2: Web Tool for predicting mental illness

The web tool creation, the main proposed system gets input from the user processes it and it gives the

prediction for mental illness.

V. CONCLUSION AND FUTURE WORK

These days, mental health is a very delicate and significant subject. It is essential to leading a balanced and healthy life. Emotions, behavior, and thoughts are all impacted by mental health. It may have an impact on one's efficiency and productivity in a person. A WHO study predicts that depression will be the primary cause of mental illness worldwide, and that people must prioritize their mental health in order to lead healthy social and professional life. Those who are uncomfortable consulting a human for a diagnosis can use online predictors to get their results.

We have first encoded the data in order to perform the prediction. We developed a system in the future that can identify a particular mental illness a person has. Based on how they answer the questions on our website, the customer gets recommendations and a likelihood of their mental health condition. Based on the accuracy we were able to achieve, we can infer that the output displays the accurate result and that there is little chance of the sickness being misclassified.
Developing a system in the future that can identify a particular mental illness a person has, but it requires a lot of data to be gathered.

VI. REFERENCES

[1]. Vidit Laijawala, Aadesh Aachaliya, Hardik Jatta and Vijaya Pinjarkar," Classification Algorithm based mental health prediction using data mining," 2020

[2]. Nor Safika Mohd Shafiee and Sofianita Mutalib," Prediction of Mental Health Problems among Higher Education Student Using Machine Learning," 2020

[3]. Rohizah ABD Rahman, Khairuddin Omar,
Shahrul Azman Mohd Noah, Mohd Shahrul Nizam Mohd Danuri And Mohammed Ali Al-Garadi," Application of Machine Learning Methods in Mental Health Detection: A Systematic Review,"2020

[4]. Juan Li," Analysis of the Mental Health of Urban Migrant Children Based on Cloud Computing and Data Mining Algorithm Models," 2021

[5]. Muhammad Shahbaz, Shahzad Ali, Aziz Guergachi,Aneeta Niazi and Amina Umer," Classification of Alzheimer's Disease using Machine Learning Techniques,"2019

[6]. Chang Su, Zhenxing Xu, Jyotishman Pathak and Fei Wang," Deep learning in mental health outcome research: a scoping review,"2020

[7]. Susel Góngora Alonso, Isabel de la Torre-Díez, Sofiane Hamrioui, Miguel López-Coronado, Diego Calvo Barreno,Lola Morón Nozaleda, Manuel Franco," Data Mining Algorithms and Techniques in Mental Health: A Systematic Review",2018

[8]. Subhan Tariq, Nadeem Akhtar, Humaira Afzal, Shahzad Khalid, Muhammad Rafiq Mufti, Shahid Hussain, Asad Habib, And Ghufran Ahmad," A Novel Co-Training-Based Approach for the Classification of Mental Illnesses Using Social Media Posts", 2019.

[9]. J. Ruiz de Miras, A.J. Ibanez-Molina, M.F. Soriano, S. Iglesias-Parro "Schizophrenia classification using machine learning on resting state EEG signal ", (2023).

[10]. Tianlin Zhang, Kailai Yang, Shaoxiong Ji, Sophia Ananiadou "Emotion fusion for mental illness detection from social media: A survey" (2023).

[11]. Shaurya Bhatnagar, Jyoti Agarwal, Ojasvi Rajeev Sharma "Detection and classification of anxiety in university students through the application of Machine Learning" (2023).