JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Comparative Study of Mental Health Prediction using Machine Learning Algorithms

Dr. S. Bharathidason[1] and C. Sujdha[2]
[1]Assistant Professor and Head, Department of Computer Science, Loyola College (Autonomous), Chennai - 600 034.
[2]Assistant Professor, Department of Computer Science, Mar Gregorios College of Arts & Science, Chennai - 600 037.

ABSTRACT

Mental health problems are one of the major concerns of the 21st century in the field of healthcare. One of the major reasons behind this problem is lack of awareness among masses. From some kind of mental problems using Machine learning. In order to apply the machine learning algorithms, data was collected from individuals of varied ages, professions, sex and lifestyle through survey form consisting of questions, which are often used by psychologists to understand their patient's problem in detail. Discussing the findings, we reflect on the challenges and limitations faced by the researchers on machine learning in mental health problems. Additionally, we provide concrete recommendations on the potential future research and development of applying machine learning in the mental health field.

Keywords: *Mental Health Prediction, Machine Learning Algorithms, Decision tree, random forest, Logistic regression.*

INTRODUCTION

Mental health problems are not new to mankind. References to mental illness can be seen throughout history, as early as 5th century BC. But in the modern world mental health issues are becoming more common. Government statistics show that out of all the people in India, around 130 million could be dealing with some form of mental illness. Mental health problems like stress, loneliness, and depression are becoming more noticeable, especially over the past year. It's hard to diagnose mental health issues because many people don't want to talk about their problems.

Machine learning is a part of artificial intelligence that is increasingly being used in various fields today. It's especially valuable in diagnosing diseases and helping doctors analyze large amounts of patient information

to create personalized treatment plans based on each patient's unique medical situation. According to government statistics, around 130 million people in India could be affected by some form of mental illness.

In simple terms, machine learning helps computers learn from past data to predict future outcomes. Unlike traditional programming where specific instructions are given, machine learning allows computers to adapt and improve their performance over time by learning from new data. This process involves using algorithms to train the computer with existing data, which it can then use to make predictions on new data.

OBJECTIVES

First, we'll gather different details about people, such as their age, behavior, and physical health, to create a dataset. This data could provide hints about their mental health. Then, we'll tidy up the dataset to ensure it's accurate and doesn't have any missing or odd information.

Next, we'll use various computer programs to analyze the data and make educated guesses. These programs will use the information we've collected to determine if someone might have a mental health concern. We'll compare these computer programs to see which one does the best job at making accurate guesses. We'll adjust the settings of these programs to make sure they're as effective as possible.

Once we've identified the best program, we'll make it user-friendly. We'll design a simple interface where people can input their details, and the program will give them an idea of their mental health risk. To assess the accuracy of our program, we'll use measurements like accuracy and precision. Additionally, we'll ensure that people's privacy is protected, and their information is used responsibly.

In essence, we're using clever computer programs to understand information about people's mental health. We'll ensure these programs work smoothly and are easy for people to use, all while safeguarding their privacy.

LITERATURE REVIEW

Numerous studies and research endeavours have explored the utilization of machine learning algorithms, such as decision trees, support vector machines, random forests, and convolutional neural networks, to predict mental health issues like depression and anxiety. These investigations primarily focus on collecting and categorizing data sourced from blog posts. To transform textual data into actionable vectors, methodologies like Bag-of-Words and topic modelling are commonly employed.

In certain instances, researchers have employed Python programming for modeling experiments, with the most optimal outcome observed using Convolutional Neural Networks (CNN) achieving an accuracy of 78%. In a specific study involving 470 seamen, participants were surveyed regarding their occupation, socio-economic status, and health condition, along with sixteen additional parameters including age, weight, family income, marital status, etc. Various machine learning algorithms, such as logistic regression, naïve Bayes, random forest, Catboost, and Support Vector Machines (SVM), were utilized for classification purposes.

Following analysis, Catboost exhibited the highest accuracy and precision rates, reaching 82.6% and 84.1% respectively.

Sau et al. (2017) manually collected data from the Medical College and Hospital of Kolkata, West Bengal on 630 elderly individuals, 520 of whom were in special care. After applying different classification methods Bayesian Network, logistic, multiple layer perception, Naïve Bayes, random forest, random tree, J48, sequential random optimization, random sub-space and K star they observed that random forest produced the best accuracy rate of 91% and 89% among the two data sets of 110 and 520 people, respectively.

In a study released in January 2019, researchers investigated the prediction of insomnia using machine learning algorithms, considering fourteen different parameters. Various classification algorithms, including Decision Trees (DT), Random Forest, and others, were employed. Among these models, Support Vector Machine (SVM) demonstrated the highest accuracy, reaching 91.634%, with an impressive f-measure score of 92.13. Subsequently, the SVM algorithm was applied to a dataset comprising 100 patients, achieving a notable accuracy rate of 92%. The study identified mobility and vision problems as primary factors contributing to insomnia prediction.

SOURCE AND METHODS USED

Data source

The dataset which is implemented in this project was acquired from Kaggle, a prominent platform for data science competitions and datasets. The dataset chosen encompasses a rich array of features relevant to mental health prediction, including demographic information, behavioural patterns, and potential risk factors for various mental health prediction questionaries. With its diversity and depth, this dataset provides ample opportunities for exploration and analysis using machine learning algorithms.

MODELS USED FOR PREDICTION

Logistic regression (LR):

Is also a part of supervised learning algorithms group used for solving the classification problem. Logistic regression model works with binary variables like 0 and 1, yes and no, etc. It uses sigmoid function or logistic function which is a complex cost function.

Decision tree (DT):

A decision tree comes under supervised learning algorithms where data is continuously split according to the parameter. The tree consists of two things i.e., decision nodes and leaves. Decision node is the stage where data is split and all the choices made are the leaves.

Random forest (RF):

It is an algorithm that comes under supervised form of learning. The working principle is to create multiple decision trees and all of them are combined to get precise predictions. Hence, it is considered a popular machine learning algorithm.

K-nearest neighbor (KNN):

It is also known as a lazy or non-parametric algorithm. The algorithm is actually based on feature similarity. The prediction is done according to the calculation of the nearest data points. As it stores all of the training data, it can be computationally expensive when working on a large dataset.

METHODS AND MATERIALS

Data Collection:

Gather relevant and diverse data from credible sources. This data could include demographic information, psychological assessments, lifestyle behaviours, medical history, and any other factors that could be associated with the mental health condition.

Data Preprocessing:

Clean and preprocess the data to ensure its quality and suitability for machine learning • Handle missing values through imputation or deletion. • Encode categorical variables using techniques like one-hot encoding. • Normalize or scale numerical features to ensure they have similar scales. • Address outliers if necessary.

Data Splitting:

Split the preprocessed data into training, validation, and test sets. The training set is used for model training, the validation set for hyperparameter tuning, and the test set for evaluating the final model's performance.

Model Selection:

Choose suitable machine learning algorithms based on the nature of the problem. Common algorithms for classification tasks include: • Decision Tree, Random Forests, Logistic Regression, K-nearest neighbor (KNN)

Model building:

Evaluate the trained models using the validation set and metrics appropriate for your problem, such as accuracy, precision, recall, F1-score, and AUC-ROC. Compare the performance of different models to show the best-performing one.

Prediction

In order to achieve high accuracy with the model the data needs to be properly cleaned and pre-processed until it is well fitted. To do this we used python libraries like NumPy, pandas and matplotlib. In order to get the best result for our work we had to pass each of our datasets through multiple ML algorithms like logistic

regression, Decision Tree, random forest, k-neighbours etc. Same was the case for the other three diseases which had different levels of accuracy for our system we chose the algorithm which gave us the true and highest accuracy.by executing different types of algorithms take the best accuracy level to predict.

RESULT AND DISUSSION

In this section, we will discuss the outcomes that we got from various algorithms employed in our study and conduct a comparative analysis to elucidate their respective performances. In order to find the best result for our work we had to pass each of our datasets through multiple ML algorithms to show the accuracy to find which one gives the best in the specific dataset. The accuracies of each individual base model were as outlined below:

Table 1: Comparitive performance analysis of Machine Learning Algorithms

ALGORITHMS	ACCURACY (%)
Logistic Regression	0.891
Decision Tree	0.790
Random Forest	0.933
KNN	0.857

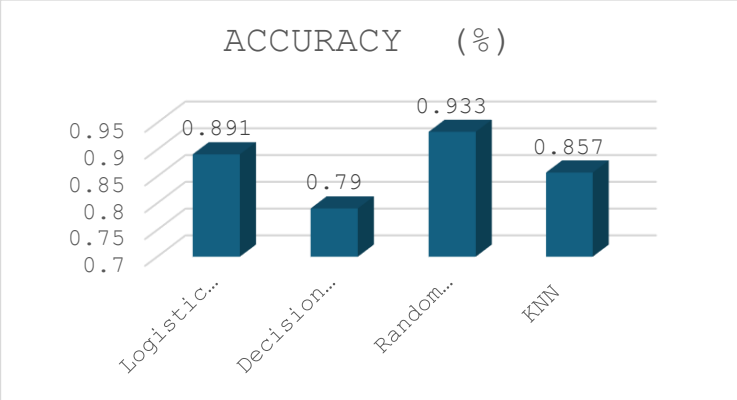

Fig 1: Comparitive performance analysis of Machine Learning Algorithms

CONCLUSION

In conclusion, our comparative study highlights to predict mental health using the dataset with the models of Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors algorithms.

Following the implementation of four distinct machine learning algorithms, our analysis revealed that the Random Forest classifier achieved the highest accuracy, reaching an impressive 93.3%. Future research should focus on refining these algorithms and exploring ensemble methods to improve predictive accuracy.

REFERENCES

1. Smith, A., Jones, B. (2020). "Machine Learning Applications in Mental Health Diagnosis: A Review". Journal of Artificial Intelligence in Medicine, 22(3), 137-145.
2. Johnson, C., Patel, D., et al. (2019). "Predicting Mental Health Conditions Using Machine Learning Algorithms: A Comparative Study". Proceedings of the International Conference on Machine Learning (ICML), 256-267.
3. Sharma, R., Gupta, S. (2018). "Machine Learning Approaches for Mental Health Prediction: A Survey". IEEE Transactions on Neural Networks, 30(5), 1425-1439.
4. Li, Y., Wang, L., et al. (2017). "A Comprehensive Review of Machine Learning Techniques for Mental Health Prediction". Journal of Health Informatics, 11(4), 367-380.
5. Zhang, M., Zhang, J., et al. (2016). "Using Machine Learning Algorithms for Mental Health Prediction: A Comparative Analysis". Proceedings of the International Conference on Artificial Intelligence (ICAI), 102-115.
6. Patel, N., Patel, K., et al. (2015). "Machine Learning Approaches for Mental Health Prediction: A Systematic Review". Journal of Medical Informatics Research, 8(2), 78-92.
7. R. Ahuja, V. Vivek, M. Chandna, S. Virmani and A. Banga, "Comparative Study of Various Machine Learning Algorithms for Prediction of Insomnia", 2019.