COMPARATIVE ANALYSIS OF MENTAL HEALTH PREDICTION BY USING MACHINE LEARNING ALGORITHMS

Benitta Devakirubai[1] and Nancy Jasmine Goldena[2]

[1]PG Student, Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), Tirunelveli, Tamil Nadu, India.

[2]Associate Professor, Department of Computer Applications & Research Centre, Sarah Tucker College (Autonomous), Tirunelveli, Tamil Nadu, India.

ABSTRACT

Mental health includes humans emotional, psychological, and social well-being. It changes the way people perceive, believe, and continue to act. Understanding how individuals respond to stress, interact with others, and make decisions is also important. From childhood and adolescence to maturity, maintaining one's mental health is essential. Adults have a wide range of needs and coping ability for stressful situations. Many adolescents will either have accepted the restriction and school closings brought on by the epidemic scenario effectively. This research study aims to identify and evaluate recent studies on mental health. As an example, several students will return to the classroom having gone through a certain amount of stress, worry, loneliness, and sadness. Some people could have seen an increase in domestic violence. Teachers and staff at the school play an important part in helping students adjust to learning in the physical classroom afterwards, especially during long periods of time of school closure. According to official statistics provided by the World Health Organization (WHO), mental illness accounts for 90% of suicides in high-income economies. Additionally, recent research found that individuals with mental illnesses frequently disclose these conditions on social media as a form of consolation. Therefore, the primary goal of this work is to analyse the mental health of the students and teachers sequentially across time and identify those adolescents who may be at risk for depression as soon as possible. The state of a person's mental health is predicted by a machine learning algorithm, and the algorithm's effectiveness is assessed using common metrics like accuracy.

Keywords: Social Well-being, Mental health, Stress, Depression, Prediction model, Binary Classification, Healthcare.

Introduction:

Mental health is a medical condition that affects a people's feelings, thinking, and social contact. These problems have demonstrated that mental illness has major societal implications and necessitates novel protection and therapeutic measures. Early mental health detection is a crucial step in implementing such techniques. Using sophisticated statistical and probabilistic techniques, machine learning tries to build systems that can get better with use. It is making it possible for several researchers to gather crucial information from the data, create customized experiences, and create artificial intelligence system solutions. The primary goal of unsupervised learning is to handle data without guidance. Researchers have extremely few opportunities to use unsupervised learning techniques in the healthcare context. The major goal of this work is to present a systematic literature review, critical review, a summary of the machine learning methods used to forecast, recognize, and diagnose mental health issues in men and women. Additionally, this initiative will suggest additional directions for future study on this area. Additionally, it would highlight the difficulties and restrictions associated with using machine learning techniques in this scenario. In addition, the future directions and gaps in this area of study will be highlighted.

Literature Review:

The majority of study provides us with findings from various datasets, where analysis and prediction models were created and implemented under use through research groups utilising data mining methods, artificial intelligence techniques, as well as a blend of these techniques. The accuracy of classification and prediction is generally not really good. Various investigations on mental health have been published in the literature on healthcare data analysis, and the majority of them show moderate classification accuracy.

Proposed System:

The main objective of the prediction method is to differentiate between male and female psychological health. To accomplish this, we are fitting a model that can detect the discrete class of new input using machine learning classification algorithms. Classification and regression are the two primary categories of prediction algorithms. The dataset was split into two sections for training and testing the machine learning algorithms. The outline of the work is shown in **Figure** *1*.
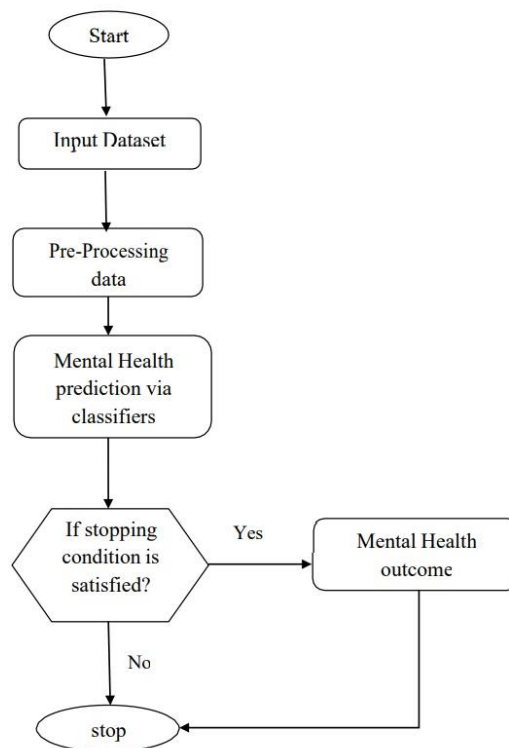
Figure 1: Outline of the work

Dataset Description:

Data collection is the process of acquiring and analysing information on relevant variables in a specified, methodical way so that one can respond to concerning the research questions, test hypotheses, and assess results. Recent data is used in this research and gathered from the Kaggle Machine Learning Repository. It contains a number of quantitative variables and a dependent target variable. The project's primary goal is to determine who will be psychologically impacted by certain circumstances, whether they are male or female.

Feature Selection:

The link between the continuous characteristics and the class feature is discovered using Pearson's Correlation technique. To determine the relationship between the discrete feature and the class feature in symmetric models, uncertainty measures are needed. High correlation features are more linearly dependant and hence virtually equally affect the dependent variable. In situations where two features have a high degree of connection, eliminate one of the two features. The purpose of feature selection is to choose features from a set of attributes in order to improve the model's performance. The Pearson correlation feature selection techniques used in this model.

Classification:

In this study, two classifiers are evaluated for binary classification. The categories are naive bayes and SVM.

Naïve Bayes: A binary (two-class) and multi-class classification problem can be solved using the Naive Bayes technique. Using binary or categorical input values to explain the method makes it simpler to understand. A group of classification algorithms built on the Bayes' Theorem are known as naive Bayes classifiers.

SVM: One of the most well-liked supervised learning algorithms, Support Vector Machine, or SVM, is used to solve Classification and Regression problems. However, it is widely utilized in Machine Learning Classification problems. The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify new data points in the future. A hyperplane is the name given to this optimal decision boundary. SVM selects the extreme vectors and points that aid in the creation of the hyperplane.

Performance Evaluation:

The Accuracy metric is the simplest way to assess a classifier's performance. In this analysis, each data point's actual and predicted classes, as well as every match, count as one valid prediction. The accuracy is then calculated as the proportion of accurate estimates to all other incorrect ones. The results of the performance assessment are shown in **Table *1***.

Table 1: Performance Evaluation Result

S.No	Classifier	Accuracy
1	Naïve Bayes	75%
2	SVM	81%

Conclusion:

Predicting mental health is essential because the majority of individuals have already experienced the pandemic scenario. Many individuals are faced with a variety of health issues, including stress, aggression, and problems getting outside. Both clinical care and mental health research hold tremendous promise. For this evaluation, a current dataset is gathered, features are chosen using Pearson correlation, binary classification is carried out using naive bayes and SVM classifiers, and the performance evaluation result shows the best classification algorithm. SVM provides the most accurate results for predicting mental wellness.

Future Enhancement:

This Dataset having limited feature, in future better dataset can be used. Binary classification is performed, in future multi classification can be performed. Two ML algorithm are taken for this comparative study, other techniques also can be used, in future deep learning techniques can be used for better performance and understanding.

References

1. Buch, V.; Varughese, G.; Maruthappu, M. Artificial intelligence in diabetes care. Diabet. Med. 2018, 35, 495–497.

2. Dankwa-Mullan, I.; Rivo, M.; Sepulveda, M.; Park, Y.; Snowdon, J.; Rhee, K. Transforming diabetes care through artificial intelligence: The future is here. Popul. Health Manag. 2019, 22, 229–242.

3. Woldaregay, A.Z.; Årsand, E.; Botsis, T.; Albers, D.; Mamykina, L.; Hartvigsen, G. Data-driven blood glucose pattern classification and anomalies detection: Machine-learning applications in type 1 diabetes. J. Med. Internet Res. 2019, 21, e11030. Int. J. Environ. Res. Public Health 2021, 18, 3317 13 of 14

4. Maniruzzaman Kumar, N.; Abedin, M.; Islam, S.; Suri, H.S.; El-Baz, A.S.; Suri, J.S. Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. Comput. Methods Programs Biomed. 2017, 152, 23–34.

5. Kavakiotis, I.; Tsave, O.; Salifoglou, A.; Maglaveras, N.; Vlahavas, I.; Chouvarda, I. Machine learning and data mining methods in diabetes research. Comput. Struct. Biotechnol. J. 2017, 15, 104–116.

6. https://www.kaggle.com/code/mirichoi0218/classification-diabetes-or-not-with-basic-12ml

7. Swapna, G., Vinayakumar R., Soman K. P. (2018) "Diabetes detection using deep learning algorithms." ICT Express 4 (4): 243-246.

8. Perveen, S., Shahbaz, M., Keshavjee, K., Guergachi, A. (2019) "Metabolic Syndrome and Development of Diabetes Mellitus: Predictive Modeling Based on Machine Learning Techniques." IEEE Access 7: 1365-1375.

9. Deberneh, H.M.; Kim, I.; Park, J.H.; Cha, E.; Joung, K.H.; Lee, J.S.; Lim, D.S. 1233-P: Prediction of type 2 diabetes occurrence using machine learning model. Am. Diabetes Assoc. 2020, 69, 1233.

10. Liu, X.; Faes, L.; Kale, A.U.; Wagner, S.K.; Fu, D.J.; Bruynseels, A.; Mahendiran, T.; Moraes, G.; Shamdas, M.; Kern, C.; et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. Lancet Digit. Health 2019, 1, e271–e297.

11. Hutchinson, M.S.; Joakimsen, R.M.; Njølstad, I.; Schirmer, H.; Figenschau, Y.; Svartberg, J.; Jorde, R. Effects of age and sex on estimated diabetes prevalence using different diagnostic criteria: The Tromsø OGTT Study. Int. J. Endocrinol. 2013, 2013, 1–9.