

PSYCHOMETRIC CHECKUP

Dr. R. B. Joshi^[1], Sarvesh Patil^[2], Akshay Chaudhari^[3], Tarunteja Bhupatwar^[4], Mayur Wagh^[5]

JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune-411033

Department of Information Technology

ABSTRACT:

In human life, mental and emotional states play an extremely prominent role. Nowadays, mental health is just as important as physical health. Due to a hectic lifestyle and an unbalanced work environment, many people are suffering from various mental health issues.

Proper treatment and counseling are required for those who are suffering from depression, stress, and anxiety. It is especially important to accurately diagnose the mental illness of a patient because even a minor mistake can directly harm the patient's health. Mental health plays a significant role in every phase of life, from childhood to adulthood.

According to the World Health Organization (WHO), more than 50% of working people are suffering from different mental health situations. Therefore, it is time to focus on the mental health of individuals. This paper focuses on the reasons for the worst mental illnesses and how to identify these problems. Based on various machine learning techniques and feature extraction methods, we can identify the correct detection of mental illnesses.

Different machine algorithms, such as Support Vector Machine, Random Forest, and Decision Tree, have different accuracies on different datasets. Based on the available data, this article presents a critical appraisal analysis of mental health detection.

KEYWORDS : *Mental health, Depression detection, Machine learning, Logistic regression, SVM, Random Forest.*

1.1 INTRODUCTION

Depression can affect people of all ages and can be very risky, causing problems such as anxiety attacks, death after a heart attack, high blood pressure, and diabetes. Therefore, it is important to recognize its symptoms and find the underlying cause to provide proper treatment.

The World Health Organization (WHO) defines "mental state" as the ability to cope with the stress of life according to one's ability, to function normally and productively, and to contribute to

mankind [1]. Factors that can affect both physical and mental health generally result from a person's lifestyle, including a deteriorating financial situation, work stress, family problems, relationship problems, violence, and environmental factors [2]. Approximately 450 million people worldwide suffer from mental illness, accounting for 13% of the global burden of disease.[3]. The WHO also estimates that one in four women suffer from a mental disorder at any stage of their lives [4].

Psychological health issues must be identified and addressed at an early stage. Early detection,

effective treatment, and accurate recognition can reduce the number of people suffering from mental health problems [5]. Conventional methods of identifying psychological illnesses usually involve self-reports, face-to-face interviews, or questionnaires. However, these methods are usually labor-intensive and time-consuming [6]. Previous research has used technologies such as smartphones and wearable sensors in medical management and psychiatric disease detection. However, these techniques are typically used by those who have already been diagnosed with a psychological illness and have been observed for a period of time [10].

1.2 LITERATURE REVIEW

1) Vidit Lajiwala, Aadesh Aachaliya, Hardik Jatta, Vijaya Pinjarkar

In this paper the algorithms used by these people are Decision Tree, Random Forest Tree, Logistic Regression, Naive Bayes through which they have achieved 82.2%, 79.3%, 81.4%, 78.7% respectively.

2) U Sairam, Santosh Voruganti

The system proposed by authors based on MLP classifier, Support Vector Machine, Decision Tree algorithms used in this paper and they have achieved 65.93, 68.89%, 71.85% accuracy respectively.

3) Anu Priya, Shruti Garg, Neha Prerna Tigga

Algorithms used by authors are Decision Tree, Support Vector Machine (SVM), Random Forest tree, Naive Bayes, KNN. The accuracy achieved using these algorithms were varied for Anxiety, Depression, Stress. The average accuracy achieved for anxiety is 73.3%, for depression is 77.8%, and for Stress is 62.8%.

4) Ashley Tate, Ryan C.McCabe, Henrick Larsson, Sebastian Lundstrom, Paul Lichtenstein

Algorithms used are Random Forest and SVM . The model performance determined by the AUC curve. They achieved accuracy of 95% by random forest and same for support of the machine.

5) Sofinata Mutalib, Nor Safika Mohd Shafiee, Shuzlina Abdul-Rahman.

Author used different algorithms to test their model. Algorithms used are Decision Tree, MLP, Naive Bayes. They achieved accuracy around 84.44%, 80%, 74.81% respectively.

6) Rohizah ABD Rahman, Khairuddin Omar (Member IEEE)

The analysis in this paper consists of data sources, feature extraction methods, and classifier performance in machine learning or deep learning techniques. We also examine pre-mental health detection by identifying data analysis methods, challenges, and limitations.

7) Pavan Kumar Reddy Yannam, Vineet Venkatesh.

In this ARU Stress Detection System has been used. They conducted a research study with sixty students which measured the impact of providing three different patterns of tailored explanations (belief-based, goal-based, and belief and goal-based explanation) on the students' intentions to change the recommended behaviors and the relationship built with the ECA.

8) Xiaofeng Wang, Hu Li, Chuanyong Sun

In this paper different algorithms were used. After doing many experiments to compare algorithms, it was found that, Feature selection algorithms combined with BPNN can improve prediction accuracy. Furthermore, the GCBA-BPNN-4 model is suitable for predicting the mental health healthcare workers at public health events.

9) Jetli Chung and Jason Teo

In this review paper a total of 20 to 30 research articles were included. This paper has minimum error rates compared with other methods. The accuracy of 73.7% is received after the classification

10) Manju Lata Joshia, Nehal Kanoongo

Depression can be detected through social media sites like Facebook, twitter, LinkedIn, etc. On the regular basis whatever tweets have been done, on the basis of that tweets level of depression has been checked by applying different algorithms. So, the Multinomial Naive Bayes algorithm worked better

as compared to the SVM in terms of accuracy and F1 score.

1.3 EXISTING SYSTEMS

1) Coached mobile app platform:

In 2020, Graham-A.K. published a paper that explores the potential of smartphone platforms for addressing mental illness. Given the widespread use of mobile devices among adults, mobile apps can provide a convenient and effective tool for delivering treatment. With the ability to track and record treatment procedures through the app, patients can benefit from remote monitoring and reduced clinic visits. This presents a promising approach for improving access to mental health services and promoting self-management of mental illness through mobile technology.

2) Mental illness stability:

In a 2018 study conducted by Gustavson K, it was demonstrated that the severity of mental disorders tends to escalate over time, and their chronic treatments can have a significant impact. However, early detection and symptom identification can lower the risk of complications. To detect early symptoms, DAAS questionnaires can be utilized.

3) Self monitoring and emotional well being :

In a 2018 publication by Bakker D, the topic of self-monitoring and emotional well-being was explored. The study highlighted the importance of real-time detection of mood swings as a means of identifying potential illness. The use of daily questionnaires was found to be an effective method of detecting such changes. Additionally, mobile applications can be utilized to detect suspicious responses and alert the user of potential issues. By implementing these strategies, individuals can proactively monitor their emotional well-being and take appropriate action when necessary.

4) Common Mental disorders:

Specifically, Steel and Marnane (1980) highlight that individuals within the working age group, which spans from 24 to 40 years old, face the highest risk of developing mental disorders due to increased stress and reduced opportunities for

relaxation. Despite depression, hypertension, and anxiety being prevalent mental health issues within this group, many individuals dismiss or misinterpret their symptoms due to misconceptions surrounding mental health. The authors also note that untreated mental disorders can lead to chronic conditions, further emphasizing the importance of early diagnosis and intervention.

5) Mental Health apps :

The paper authored by Kenny R-Dooley in 2016 discusses the manifestation of depression symptoms and behaviors during the adolescent phase. To support this claim, the paper included the results of a survey conducted on a sample of 10,000 individuals aged between 12 and 18. The accuracy of the CNN meta-algorithms utilized in the study yielded a remarkable 92.79% accuracy. Subsequently, the findings of the study were incorporated into a mobile application.

1.4 TECHNIQUES

1) Logistic Regression:

Logistic regression is a widely used supervised learning algorithm that is employed to predict categorical dependent variables based on a set of independent variables. It is primarily used in classification tasks, and the output of the model is a probabilistic value that falls between 0 and 1, as opposed to an exact categorical value. The output values could be true or false, yes or no, or 0 or 1. The main difference between linear regression and logistic regression is that linear regression is used to solve regression problems, whereas logistic regression is used to solve classification problems.

2) Random Forest Algorithm (RF):

The Random Forest algorithm is built upon the decision tree algorithm, which partitions data records into smaller features. This approach is known for its simplicity and straightforward calculations. The Random Forest algorithm uses a collection of decision trees that are trained on random subsets of data, ultimately leading to improved accuracy in predicting outcomes.

3) K Nearest Neighbor (KNN):

The K-Nearest Neighbor (KNN) algorithm is a powerful pattern recognition technique utilized for predicting the class or continuous value of a new data point by analyzing the K closest neighbors (data points) within the training set. This algorithm is particularly advantageous for classifying objects based on various features. The KNN algorithm examines the distances between the new data point and each of the K nearest neighbors, thus enabling the determination of the class or value of the new data point. This approach is efficient for handling non-linear decision boundaries, and it is capable of performing both classification and regression tasks. In essence, the KNN algorithm is a versatile and reliable tool for pattern recognition and data analysis.

4) Naive Bayes (NB):

The Naive Bayes algorithm is a supervised learning technique that utilizes Bayes' theorem to address classification problems. It is frequently employed in text classification tasks that involve vast amounts of data. The Naive Bayes Classifier is widely regarded as one of the most efficient and straightforward classification algorithms, allowing for the rapid construction of machine learning models capable of making swift predictions. As a probabilistic classifier, the Naive Bayes algorithm operates on the basis of object probability, enabling it to make informed predictions. Its applications are diverse, including spam filtering, sentiment analysis, and article classification.

1.5 METHODOLOGY

1) Data Cleaning: It is crucial to ensure that the dataset used for the implementation is of high quality. In order to achieve this, we obtained a well-balanced and reliable dataset from Kaggle. To create a clean and dependable set for training, validation, and testing purposes, we meticulously sorted through any residual duplicates and missing data. This allowed us to have a robust dataset that can be utilized for our experiments with confidence. We have taken all necessary steps to ensure the quality and reliability of the dataset, as it is a crucial aspect of our research.

2) Data Encoding: This is a data modification process which involves assigning weightages to

categorical responses. Typically, the data is encoded into binary strings of 1's and 0's for future classification. This encoding allows for efficient analysis and can aid in identifying patterns within the data.

3) Feature Scaling: Ensuring standardization of independent features in a dataset is crucial as it enables them to be measured on a fixed scale. Neglecting this step may lead to biased results in machine learning algorithms, as they tend to assign higher importance to features with larger magnitudes, regardless of the units in which they are measured. Therefore, it is imperative to apply standardization techniques to ensure a fair and unbiased analysis.

4) Tuning: Optimizing data is a crucial step in developing an accurate model without the risk of overfitting or high variance. It is vital to establish a reliable model. Tuning the data involves adjusting the parameters of the model to improve its performance, which ensures the accuracy of the predictions. This process also helps to eliminate any inconsistencies in the data, which can significantly affect the model's efficiency. Consequently, it is paramount to fine-tune the data to increase the model's robustness and establish trust in its outputs.

5) Decision Tree Classifier: The utilization of decision trees is driven by the necessity for flexible models capable of learning functions based on if-then-else decision rules. This technique enables the extraction of knowledge from training data regarding target variable behavior, which can be used to make predictions and classifications during testing. Decision trees are an effective tool for enhancing model flexibility and allowing for more accurate predictions and classifications.

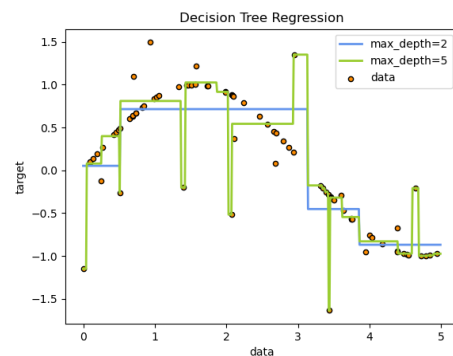


Fig. 1. Decision Tree Regression

6) Random Forest: The Random Forest classifier algorithm is widely utilized due to its simplicity and capability to make decisions on multiple subsets of the dataset, which enhances its accuracy by aggregating different compositions. This approach is an ensemble of several decision tree algorithms, enabling it to handle complex datasets with high dimensions. The Random Forest classifier is a popular choice in various fields, such as finance, healthcare, and marketing, due to its robustness and ability to provide reliable predictions. Its ability to handle missing data and deal with noise makes it a powerful tool for data analysis.

7) Bagging: Ensemble meta-algorithm of bagging is a crucial technique in machine learning that possesses exceptional capabilities in identifying patterns within statistical problems. The stability and accuracy of the bagging model make it a valuable tool for making regression and classification predictions. Moreover, bagging facilitates the reduction of overfitting and the improvement of model generalization. Its versatility allows for the aggregation of various weak learners to create a strong, robust model. In summary, bagging is a fundamental machine learning technique that improves model performance and can be applied to various prediction tasks.

8) Boosting: The algorithm serves as a valuable tool for assisting slower learning algorithms in achieving rapid and precise mastery. Its function is primarily supportive in nature, aiding those with a more gradual pace of learning to keep pace with their more swift counterparts. To accomplish this, the algorithm employs a range of techniques, such as utilizing synonyms and employing professional language, in order to enhance its efficacy. Furthermore, if necessary, additional points can be incorporated or extraneous material removed in order to optimize the algorithm's overall performance.

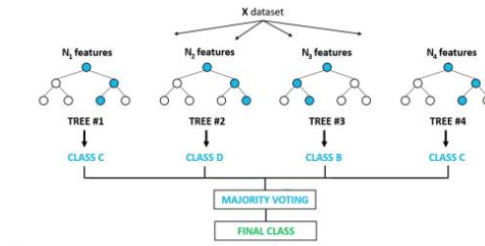


Fig. 2. Boosting Algorithm

2. MODEL EVALUATION

1) Confusion Matrix: In some cases, even the most suitable classification model can yield inaccurate results, which can be misleading. To prevent such inaccuracies, we use a confusion matrix, which provides a comprehensive overview of how well the model performed on a given dataset. By utilizing this matrix, we can gain insights into the model's strengths and weaknesses, allowing us to make necessary adjustments to improve its performance.

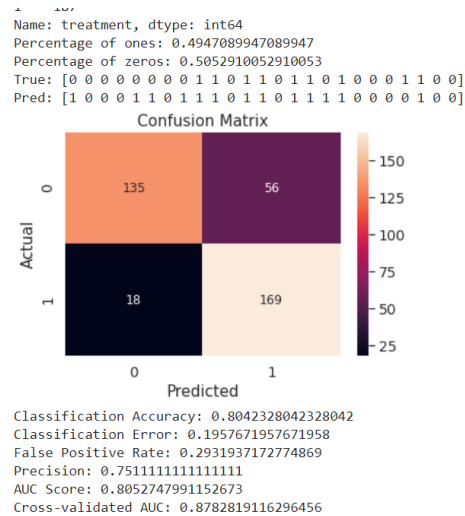


Fig. 3. Confusion Matrix

index	prediction	expected	
0	929	0	0
1	901	1	1
2	579	1	1
3	367	1	1
4	615	0	1
5	981	1	1
6	384	1	1
7	200	1	1
8	503	0	0
9	65	1	0

Fig. 4. Prediction VS Expected

2) ROC Curve : The Receiver Operating Characteristic (ROC) curve is a graphical representation that depicts the performance of a model in terms of accuracy and reliability. The maximum value of the ROC curve is 1, which corresponds to 100% accuracy. The area under the ROC curve reflects the accuracy of the model. Therefore, the ROC curve is an essential tool for evaluating the performance of a model and determining its reliability.

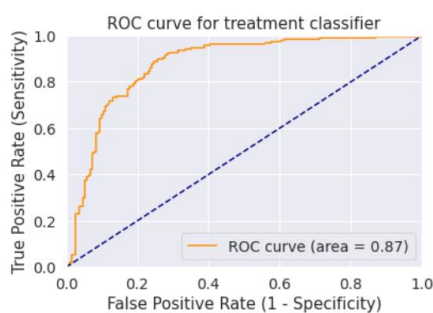


Fig. 5. ROC Curve

3. RESULTS AND DISCUSSIONS

After successfully implementing our model, we conducted a random sampling of 10 instances from the testing dataset. Our analysis revealed varying

levels of accuracy across different algorithms, as illustrated below. To ensure the reliability of our findings, we validated the accuracy measures using multiple methods, including a confusion matrix, classification report, and ROC curve.

Algorithm	Accuracy
Logistic Regression	79%
Random Forest	81%
KNN	80%
Naive Bayes	84%

Table 1. Result

4. CONCLUSION

In this research paper, we employed a range of machine learning techniques to evaluate an individual's mental state and verify their mental well-being. Our analysis involved utilizing various algorithms, including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), among others. By evaluating the efficacy of these techniques, we were able to determine the most effective algorithm and achieve improved outcomes.

An individual's mental state can provide insight into their psychological health, including conditions such as depression, anxiety, and other disorders. Detecting these conditions early on is crucial, and our machine learning-based application can aid in the identification of these issues. Additionally, it can help to address work-related stress, anxiety triggers, mood fluctuations, and negative thought patterns. Furthermore, our application can provide prognostications based on a patient's medical history.

5. REFERENCES

- [1] Sau, A., Bhakta, I. (2017)"Predicting anxiety and depression in elderly patients using machine learning technology. "Healthcare Technology Letters 4 (6): 238-43.
- [2] Tyshchenko, Y. (2018)"Depression and anxiety detection from blog posts data." Nature Precis. Sci., Inst. Computer. Sci., Univ. Tartu, Tartu, Estonia.
- [3] Saha, B., Nguyen, T., Phung, D., Venkatesh, S. (2016) "A framework for classifying online mental health-interest in depression. "IEEE journal of biomedical and health informatics 20 (4): 1008-1015.
- [4] Reece, A. G., Reagan, A. J., Lix, K. L. M., Dodds, P. S., Danforth, C. M., Langer, E. J. (2016) "Forecasting the Onset and Course of Mental Illness with Twitter Data." Scientific reports 7 (1): 13006.
- [5] Braithwaite, S. R., Giraud-Carrier, C., West, J., Barnes, M. D., Hanson, C.L. (2016) "Validating machine learning algorithms for Twitter data against established measures of suicidality." JMIR mental health 3 (2): e21.
- [6] Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., Xu, H. (2018) "Extracting psychiatric stressors for suicide from social media using deep learning." BMC medical informatics and decision making 18 (2): 43.
- [7] Al Hanai, T., Ghassemi, M. M., Glass, J.R. (2018) "Detecting Depression with Audio/Text Sequence Modeling of Interviews." Interspeech: 1716-1720.
- [8] Ramiandrisoa, F., Mothe, J., Benamara, F., Moriceau, V. (2018) "IRIT at e-Risk 2018."E- Risk workshop: 367-377.
- [9] International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue II Feb 2022
- [10] Koa Health, Barcelona, Spain. 2 Universitat Pompeu Fabra, Department of Information and Communication Technologies, Barcelona, Spain. 3 Kannact, Barcelona, Spain. 4Birmingham and Solihull Mental Health NHS Foundation Trust, Birmingham, UK. 5 University of Warwick, Warwick, UK. 6 Aston Medical School, Aston University, Aston, UK.
- [11] Sofianita Motalib¹, Nor Safika Mohd Shafiee², Shuzlina Abdul-Rahman³ ^{1,2} Faculty of Computer and Mathematical Sciences, UniversitiTeknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia. Research Initiative Group Intelligent Systems, UniversitiTeknologi MARA, 40450 Shah Alam, Selangor, Malaysia.
- [12] Ashley E. Tate^{ID1}, Ryan C. McCabe^{ID2}, Henrik Larsson^{1,3}, Sebastian Lundström^{4,5}, Paul Lichtenstein¹, Ralf Kuja-Halkola¹Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden.

