



A Real Time Analysis and Prediction of Mental Health Disorder Based on Machine Learning Technique

Karnika Shrivastava, Prashant Richhariya and Anita Soni

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

April 2, 2022

A Real Time analysis and prediction of Mental Health Disorder based on Machine Learning Technique

Karnika Shrivastava
MTECH Scholar
karnikas14@gmail.com
TIT- Advance, Bhopal

Prof. Prashant Richhariya
HOD, CSE Department
prashant1579@gmail.com
TIT- Advance, Bhopal

Dr. Anita Soni
Professor
dranita2203@gmail.com
TIT- Advance, Bhopal

ABSTRACT

Mental fitness is the aggregation of emotional, social and mental well being of a person. It impacts at the person's thinking, appearing and feeling capability. Mental fitness is a degree of managing strain and selection making with each step-in life. There is a lot records to be had that we're now capable of collect records for intellectual fitness experts via way of means of making use of this method they may benefited to clinicians the possibility to customise the professional & capable of carry out their task in higher manner in. Machine getting to know algorithms should assist decide key behavioural biomarkers to useful resource intellectual fitness experts in identifying if a affected person is vulnerable to growing a selected intellectual fitness disorder. Additionally, the algorithms might also additionally help in monitoring effectiveness of a remedy plan. This paper evaluations approximately the software of ML to intellectual fitness prediction, which incorporates a number of advantages throughout the regions of prognosis, remedy and support, research, and medical administration. With the bulk of research diagnosed focusing at the detection and prognosis of intellectual fitness conditions.

KEYWORDS: Machine learning, Appetite, Mental illness, Depression, Schizophrenia.

I. INTRODUCTION

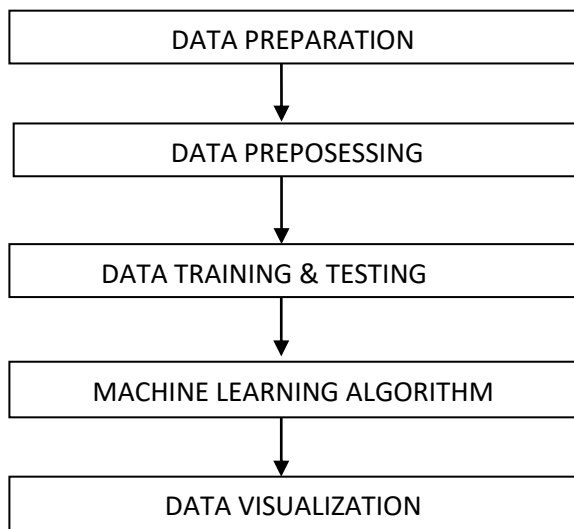
Anxiety, Depression and Schizophrenia are the maximum not unusualplace sort of intellectual disorders. Anxiety is a intellectual fitness ailment characterized via way of means of emotions of worry, worry which can be robust sufficient to intervene with one's each day activities. Depression is characterized by the way of means of consistently depressed temper or lack of hobby in activities, inflicting widespread impairment in each day life. The continual feeling of unhappiness or lack of hobby that characterises main despair can result in quite a number behavioural change. Schizophrenia impacts someone's capacity to think, experience and behave clearly. The specific reason of schizophrenia isn't always known, however a aggregate of genetics,

surroundings and changed mind chemistry and shape might also additionally play a role. This intellectual fitness ailment might also additionally encompass adjustments in sleep, appetite, strength level, concentration, each day behaviour or self-esteem. These also can be related to mind of suicide.[6] Mental contamination is an final results of imbalances in mind chemistry. The assessment of intellectual well being is extraordinarily vital to apprehend and endorse remedies for sufferers with a deviated intellectual behaviour.

This look at hired 5 tiers as follows, (1) information enter process, (2) pre-processing information, (3) KNN set of rules type process, Naive bayes and logistic regression (4) assessment of algorithms. Aim of this thesis is to expect whether or not someone could have intellectual contamination or now no longer on

the premise in their symptoms. We have amassed information from actual time remarks with the assist of google form. The dataset specially includes information of operating people and pupil age between 20 to 35years. The data consists of string attributes which we later encoded to numeric attributes for better prediction. [2]It consists of 14 questions based on which our system will predict a specific mental illness that a person suffers from mainly (Anxiety , Depression or Schizophrenia).The dataset used is in generally textual form , the data then converted into machine language for further process. Data cleaning was done manually. Firstly, it removed attributes with a lot of missing values. The missing value was data with an unfilled survey. Data cleaning consisted of five processes, namely, (1) removing attributes with a lot of missing values, (2) removing irrelevant attributes (3) removing identical attribute.Data pre-processing - Standardization is an important technique that is mostly performed as a pre-processing step before many Machine Learning models, to standardize the range of features of input data set.We have applied machine learning algorithm to create a model.

Figure 1. shows the block diagram of the implemented system. It illustrates the various modules and phases of the system



AUTHOR	ABSTRACT	METHODOLOGY	OUTCOME
Emmanuel ,Theodore ,PE pintelas	A review of machine learning prediction method for anxiety disorder	i)Naïve bayes& Bayesian network ii)ANN iii)SVM	i)95% accuracy ii)82.35% iii)78.5%
Sandhya P ,Mahekkant esaria	Prediction of mental disorder for employees in IT industry	i)Logistic regression ii)KNN iii)Decision tree iv)Random forest v)Bagging vi)Boosting vii)Neural network	79% 80.4% 80.6% 81.2% 77.7% 81.7% 81.4% Accuracy
Anu priya , shruti garg ,nehaprenat igga	Predicting Anxiety, Depression and Stress in Modern Life using Mamchine Learning Algorithm	i)Decision tree ii)Random Forest ii)Naïve bayes iv)Support vector machine	Anxiety - 0.733 Depression - 0.778 Stress-0.628 Anxiety -0.714 Depression- 0.798 Stress -0.723 Anxiety - 0.733 Depression- 0.855 Stress- 0.742 Anxiety- 0.678 Depression 0.803 Stress 0.667
Devakunchari Ramalingam, Vaibhav Sharma, Priyanka Zar	Study of Depression Analysis using Machine Learning Technique	Logistic Regression (SLR), Multilayer Perceptron Neural Networks (MLPNN), Support Vector Machine (SVM), and Random Forest (RF).	The system using this achieved average detection accuracy of 82.2% in case of males and 70.5 in case of females
V shrinivasanreddy, adityavivek, a dharun.	Machine learning techniques for stress prediction in working employees	Logistic Regression K-NN Decision Tree Random forest boosting	Boosting produced the highest accuracy

II. RELATED WORK

In [13] U. S. Reddy et al. have applied various algorithms to find the most accurate one and compared the relationship between various parameters in the dataset. In [14] M. P. Dooshima et al. have used demographic, biological, psychological and environmental factors for prediction. Different mental health experts were consulted to validate the obtained parameters. In [15] M. Srividya et al. have used 22. W. Zhang, X. Yang, S. Lui, Y. Meng, L. Yao, Y. Xiao, W. Deng, W. Zhang, and Q. Gong. 2015. Diagnostic prediction for social anxiety disorder via multivariate pattern analysis 25. Impact of COVID-19 and lockdown on mental health of children and adolescents: A narrative review with recommendations. In [7] S. G. Alonso et al. have conducted extensive review of different algorithms used for mental health prediction. Different techniques such as Association Rule Mining and Randomization were studied and their predictions were noted for our project. This paper also reviewed other algorithms such as SVM, Decision tree, KNN, ANN, Naïve Bayes.

There are different kinds of systems that currently exist. Most of them use different methodologies to predict mental illness. Some of the current systems include an online survey which predicts whether the user has mental illness or not.

Machine Learning algorithms:

Brief Introduction of the classifiers: - Naive Bayes:- This classifier belongs to the probabilistic group of classifiers in the domain of machine learning. The base of this classifier is the Bayes Theorem where the features are considered to be independent of each other. It is a very popular when it comes to classification. It is a simple model where the test (unknown) instances are assigned class tags based on the trained model. K-NN :- K-nearest neighbor model can be used as classification model or regression model. For an unclassified instance as the input we consider

the k classified instances in a constraint region and accordingly the unclassified instance is given a class whose instances are most in that region. In case $K=1$, the unclassified instance is given the class whose neighbour is nearest to it, there is no need for count as the value of k is 1.

SVM: - A Support Vector Machine (SVM) works by finding a hyper-plane that can efficiently divide the set of objects in different classes. SVM takes a labelled training data, and outputs an optimal hyper-plane which can then be used to categorize new examples. A decision plane separates set of objects having memberships of different classes. A set of training examples is provided to the algorithm and it generates a boundary in order to differentiate between the classes learning from training examples

III. IMPLEMENTED SYSTEM

Our aim in this thesis is to predict whether a person can have mental illness or not on the basis of their symptoms. we review the non-technical and technical studies dedicated to mental illness. Mental illness is an epidemic phenomenon and is generating severe harm to people, especially adults. The thesis deals with the kinds of mental illness, several work has been done in this field and also the background studies that are important for performing the analysis. Though there is barely any work trying to predict schizophrenia taking into consideration all these features, we have tried to implement such an all in one model here that will take care of all these aspects of mental illness. It illustrates the methodology proposed by us. We have gathered the data from different sites and perform the process of data acquisition and feature extraction. After the preprocessing of the data, we have them labeled as 0 (no mental illness) and 1 (mental illness) . The proposed method which we have used as three classification algorithm KNN, Naive bayes and Logistic regression The

chapter also introduces the involve features like color histogram for data visualization.

Here we have explained the setting of various parameters that has been used for performing the experiments. We have defined the proper distribution of the data as in what proportion the modalities are used in our model. Further we have analyzed our model individually for each type of modality and analyzed the results. The results are also compared by using classification algorithms like Logistic regression and KNN and observed that these two gives the best results. the accuracy we have received is 96%, which is the improvement over past research.

IV. RESULTS AND ANALYSIS

Classification using Logistic regression, naive bayes and the KNN method was carried out with different values of k. it was done to obtain the optimal k value with the best accuracy. We have used three evaluation metrics: Precision, Recall, Accuracy.

Precision: It is the ratio of data elements that are correctly classified (for both the minority and majority class) to total number of classified instances' =TP/ (TP + FP)

Table 1 Representation of Naive bayes

PREDICTION	precision	recall	f1-score
NOMENTALDISEASE	0.77	1.00	0.87
MENTAL DISEASE	1.00	0.73	0.84

Recall: The ratio of the minority class instances that are correctly classified to the

total number of actual minority class instances. $R = TP / (TP + FN)$.

F-Measure: Precision and Recall are used for performing the calculation of Fmeasure. It is calculated by taking the harmonic mean

PREDICTION	precision	recall	f1-score
NOMENTALDISEASE	0.91	0.94	0.93
MENTAL DISEASE	0.95	0.91	0.93

Table2 - Representation of Logistic Regression

of Precision & Recall. We can say that it is essentially an average between the two percentages. It really simplifies the comparison between the classifiers. $F\text{-measure} = 2 / (1/R + 1/P)$.

Table3 - Representation of KNN

PREDICTION	precision	recall	f1-score
NOMENTALDISEASE	0.91	0.92	0.96
MENTAL DISEASE	0.93	1.00	0.96

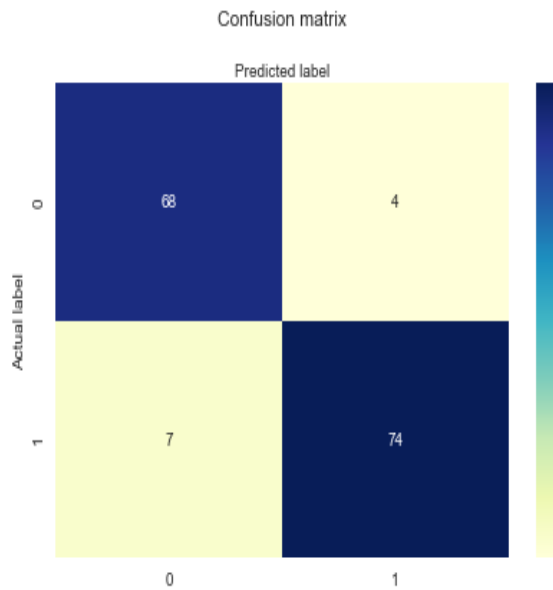


Fig 3 Confusion matrix of Logistic regression

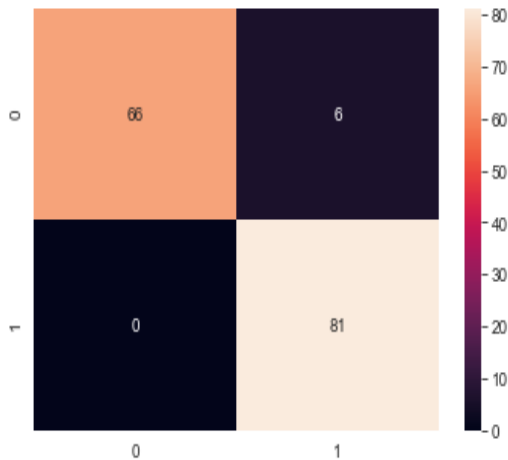


Fig 4 Confusion matrix of KNN

OUT[32]: <AxesSubplot:>

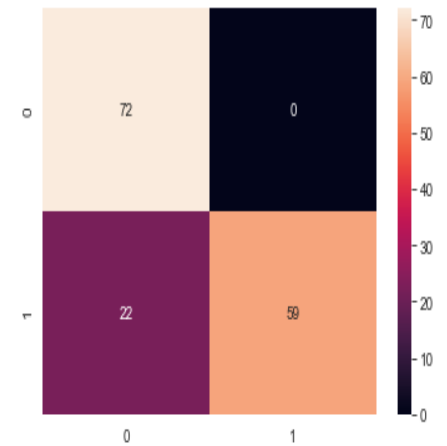


Fig 5 Confusion matrix of Naive bayes

Figures show the confusion matrix denoting the accuracy of the algorithm. It shows that 149 instances have been correctly classified as Positive while 109 have been correctly classified as negative.

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

Using the values in figures, we get the accuracy of the KNN algorithm as 0.96, Logistic regression, 0.92 and Naive bayes 0.85.

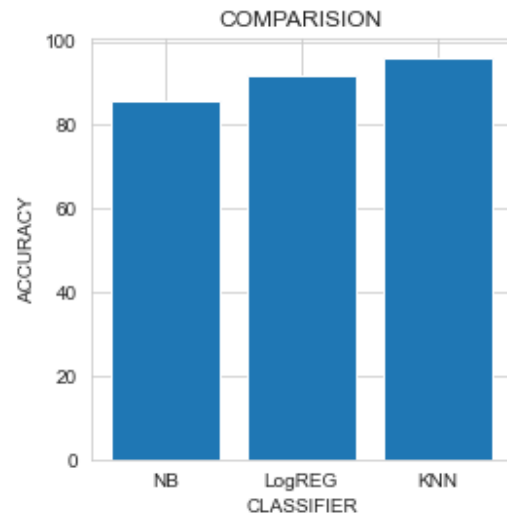


Fig 6 Comparison

V. CONCLUSION

As per the study by WHO, depression will be a major cause of mental illness in the world and people need to take more care about their mental well-being for a balanced social and professional life. People who are hesitant to approach humans for diagnosis can make use of online predictors for results.

To do the prediction, we have encoded the data first. We have then used the KNN and logistic regression algorithm and trained a model. The accuracy we received with both the algorithm was 96% and 92%. To sum up, this research obtained 70% of training data and 30% testing data which resulted in higher values of precision, recall, and accuracy compared to other data compositions.

VI. FUTURE SCOPE

In the future, we can create a system which predicts a specific mental illness by formulating a scoring system for different attributes based on their importance to create a uniform scale to measure the levels of a particular disease one can have, so that the person can be diagnosed earlier, however, extensive data collection needs to be carried out for it. Additional methods can be used to test the efficiency of the model.

VI. REFERENCES

1. DEPRESSION: A Global Crisis, WHO, https://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf March '12
2. U. S. Reddy, A. V. Thota and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 2018, pp. 1-4.
3. M. P. Dooshima, E. N. Chidozie, B. J. Ademola, O. O. Sekoni, I. P. Adebayo, "A Predictive Model for the Risk of Mental Illness in Nigeria"
4. Emmanuel, theodore, PE Pintelas [2018] "A review of machine learning prediction method for anxiety disorder"[DSAI 2018 conference Greece]. DSAI-"software development and technologies for enhancing accessibility and fighting info-exclusion".
5. Viditlajawala, Aadeshachaliya, hardikjatta, vijayapinjarkar [2020] "classification algorithm based mental health prediction using data mining" [IEEE 2020].
6. Aadityavivekthota, A dharun [2018] "machine learning techniques for stress prediction in working employee"[IEEE 2018].
7. Hakkumelmunysnah, risalatul, triyana, widiyaningtyas, felix [2018] "classification of employee mental health disorder treatment with KNN algorithm"[IEEE 2018].
8. Anu priya, Shruti garg, Neha prenatigga [2019] "predicting anxiety, depression and stress in modern life using machine learning algorithm [ICCIDS 2019]" "ICCIDS- International conference on computational intelligence and data science"
9. www.towardsdatascience.com
10. The international classification of diseases 10th edition. (referedsite)
11. www.WorldHealthOrganization.com
12. en.wikipedia.org
13. M. Srividya, M. Subramaniam and B. Natarajan, "Behavioral Modeling for Mental Health using Machine Learning Algorithms" "Journal of Medical Systems" Vol. 42(5):88 May 2018.
14. DEPRESSION: A Global Crisis, WHO,

- https://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012
15. U. S. Reddy, A. V. Thota and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)
 16. Bhattacharyya, R., & Basu, S. (2018). India Inc looks to deal with rising stress in employees. Retrieved from 'The Economic Times'
 17. Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology.
 18. Bhakta, I and Sau, A. (2016). Prediction of Depression among Senior Citizens using Machine Learning Classifiers. International Journal of Computer Applications Vol.
 19. P. N. Tan, M. Steinbach, Vipin Kumar, "Introduction to Data Mining", Pearson Education.
 20. V Krishnakumar, LathaParthiban, Alzheimer's Disease Neuroimaging Initiative, et al. 2019. A Novel Texture Extraction Technique with T1 Weighted MRI for the Classification of Alzheimer's Disease. Journal of neuroscience methods
 21. Belathur Suresh Mahanand, Sundaram Suresh, Narasimhan Sundararajan, and M Aswatha Kumar. 2012. Identification of brain regions responsible
 22. Impact of COVID-19 and lockdown on mental health of children and adolescents: A narrative review with recommendations
 23. Syaliman, K U., Nababan, E B., Sitompul O S. 2018. Improving the Accuracy of K-Nearest Neighbor Using Local Mean Based and Distance Weight. Journal of Physics : Conference Series. From chromeextension://ngpampappnmepgilofjohadhhmbhlaek/captured.html?bac k=1.