Proceedings of the Third International Conference on Inventive Research in Computing Applications (ICIRCA-2021)
IEEE Xplore Part Number: CFP21N67-ART; ISBN: 978-0-7381-4627-0

Machine Learning Techniques for Prediction of Mental Health

[1] Tarun Jain,[2]Ashish Jain,[3]Priyank Singh Hada, [4]Horesh Kumar, [5]Vivek K Verma, [6]Aayush patni

[1]tarunjainjain02@gmail.com,[2]imjainashish@gmail.com,[3]pshada1@gmail.com,[4]horeshkiet1991@gmail.com,

[5]Vermavivek123@gmail.com, [6]ayushjain826@gmail.com,

[1,2,3,5,6]SCIT, Manipal University Jaipur, [4] GNIOT Group of Institutions Greater Noida, Uttar Pradesh, India

2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) | 978-1-6654-3877-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIRCA51532.2021.9545061

Abstract— **Suicide is the 2nd leading cause of death in the world, for those aged 15-24 and about 800,000 victims of suicide yearly (all age), which is about 40 per second. Behavioural health disorder, explicitly depression, are the type of health concerns, not many are aware of. There is no way one can get treatment of something they are not aware of. So, classifying potential health disordered person is the first step towards prevention. Lifestyle is something which defines individual the best. Lifestyle including Income, age group, martial status, child, property owned, alcohol or tobacco consumption, medical expenditure, insurance or other type of investment and many more. Using 76 such kind of attributes, model will predict if the individual is victim of depression or not. The proposed model has used eight mainstream ML calculation methods, namely (Decision tree (DT), Random Forest(RF), Support Vector Machine(SVM), Naïve Bayes(NB), Logistic Regression(LR), XGBoost(XGB), Gradient Boosting Classifier(GBC) and Artificial Neural Network(ANN) to build up the expectation models utilizing a huge dataset (1429 individual's survey), bringing about precise and productive dynamics. By using various strategies and different model, this research work has attempted to get a clear and precise picture. The reason to follow various approaches is that, precise the information, work in a better way and reduce the number of suicide case. The final outcome received was 87.38 percent, which was using Support Vector Machine (SVM).**

INTRODUCTION

Mental health is one of the leading causes to suicide [10]. As most of the victims know a little about their mental health, it goes unseen. Every year more than 800,000 individuals die due to Depression or some mental health disorder and this number is not a small one or which could be ignored [5]. Most of the individuals are of the age between 15-24. Which is mostly youth and have a lot of potential to become something and give society a lot [1]. To achieve success in such cases, the potential cases must be sorted before even they know. Mental health has a lot to do with the amount earned and the way it has been spent [8]. Our spending and liabilities tell a lot about our potential mental health. examining previous cases, in which it has considered how much individual do earn and there spending habits or where there most chunk of money goes, if they use tobacco, alcohol, medical expenses, Social expenses, Education expenses, number of time they eat meat or fish, slept hungry or not, enough food for tomorrow, Doctor consulting and other kind of liabilities. Also, the family structure has been considered, in that how many members are there in family, age of the respondent, if the respondent is female, Marital status, number of children. Assets are also included, such as household size, number of farms, value of live stocks, value of durable goods, value of savings, food own production and many more. Being stress free is one of the key components for healthy life [9].

Goal was to make a ML model, using which the future possibility of mental health can be identified. Using and analysing all the data in logical manner gave us insight that there are some similarities in all these cases [2]. Using which the mental health can be predicted and using the same model, the potential mental health victim can also be predicted.

The utilized dataset has 1429 tuples and 76 attributes (Some attributes: name, surveyid, village, survey_date, femaleres, age, married, children, hhsize, edu)

The eight classification algorithms used to classify such type of mental conditions are:

Decision tree(DT),
Random Forest(RF),
Support Vector Machine(SVM),
Naïve Bayes(NB),
Logistic Regression(LR),
XGBoost(XGB),
Gradient Boosting Classifier(GBC)
Artificial Neural Network(ANN)

METHODOLOGY

The dataset has been acquired from both the individuals, who are affected and not affected with such a condition. Our approach includes utilization of arrangement systems like (Decision tree (DT), Random Forest (RF), Support Vector Machine

(SVM), Naïve Bayes (NB), Logistic Regression (LR), XGBoost (XGB), Gradient Boosting Classifier (GBC) and Artificial Neural Network (ANN).

A. Dateset

Dataset used has 76 attributed from them some are given [11]:
We have mentioned attributes with the details:

1. Surveyed : Individual Identifier
2. Village : Village Identifier
3.cons_nondurable: Non-durable expenditure (USD)
4. asset_savings : asset_savings
5. cons_alcohol : Alcohol (USD)
6. cons_tobacco : Tobacco (USD)
7. saved_mpesa: Saved money using M-Pesa
8. durable_investment: Durable Investments
9. ed_expenses : Education expenditure past month (USD)
10. early_survey : Psychology survey in 1st wave (dummy)
11. depressed: Meets epidemiological threshold for moderate depression (dummy)

B. Data Pre-processing

The process of converting some raw data into some meaningful information can be performed with some logical task and derive results [3]. This is known as data pre-processing. For making a machine learning model, this is one the most crucial step, as in this phase we decide, if the information is useful to us, or using which we can have some insights.
In this process we do clean data, that means removing null values, assigning some values, derived from similar data or mean of data [13]. In this phase we may also encounter some NA values, which we must remove, which we can do by different ways [18]. Feature scaling is used to scale different values to a same scale, so that we can perform operations easily.
There was a column which was of date, we didn't want to make our prediction based on date, or make date as a influencing factor, so we dropped the date as column. Out of 86868 there were 13196 Nan values, which we removed and applied median function.

MODEL SELECTION

There are many strategies, which can be used to find relevant results. There are two gatherings possible from which we can characterize results of those calculations, Regulated (Supervised) learning and unaided (Unsupervised) learning [14]. In Supervised learning, machine is prepared using information. Then the result is anticipated on the grounds of data or information given to system. Using information, model justifies a pattern using which the result is anticipated. They are addressed as Regression and Classification methods. Unaided learning, while we perform unaided learning, the information given to machine does not have any labelling, thus machine has to make its own ways, connecting various parameters [4]. It is used in circumstances in which we just need some connection between data or some hidden pattern which are about impossible to find using human eyes, and it is all done without any human intervention, so chances of anomaly also decreases. As per our dataset, as dependent variable or result variable we have has two possible conclusion, either the individual can be depressed or not depressed. As we had labelled dataset, we had applied classification calculation of directed learning. Eight Distinct sort of arrangement calculation of machine learning is used .
1. Decision tree (DT)
2. Random Forest (RF)
3. Support Vector Machine (SVM)
4. Naïve Bayes (NB)
5. Logistic Regression (LR)
6. XGBoost(XGB)
7. Gradient Boosting Classifier (GBC)
8. Artificial Neural Network (ANN)

A. Decision Tree (DT)

Decision tree is a concept which can be used in both, either classification or regression. In classification we mostly use decision tree, as it one of the most powerful and efficient tool [12]. Decision tree is a tree like structure classifier. After providing dataset as input in an algorithm, it makes a model out of it, further which we can use it to predict results by providing our desired values or information. The result will be anticipated on the bases of the input dataset [6]. In Decision tree, each node represents a test on an attribute, outcome of the test is represented by the branch and leaf node is treated as class label(terminal node) [17]. Decision tree is drawn up side down, which mean it starts from the top, which means it's root locates on the top and leaf at bottom.
In decision tree we have tree nodes:

1. Chance Node – represented by circle 
2. Decision Node – represented by square 
3. End Node – represented by triangle 

Construction of Decision tree
Decision tree is made on the bases of outcomes from the dependent attribute. For that we have to find entropy. Entropy is the possibility of getting a yes or no or we can say a 0 value or 1 value. To find the entropy of the class attribute, we have a formula:

$$E(S) = \sum_{i=1}^{c} -pi\,;log2\ pi$$

Here E is the Entropy, s is the set and pi is the frequentist probability of an element/class 'i.

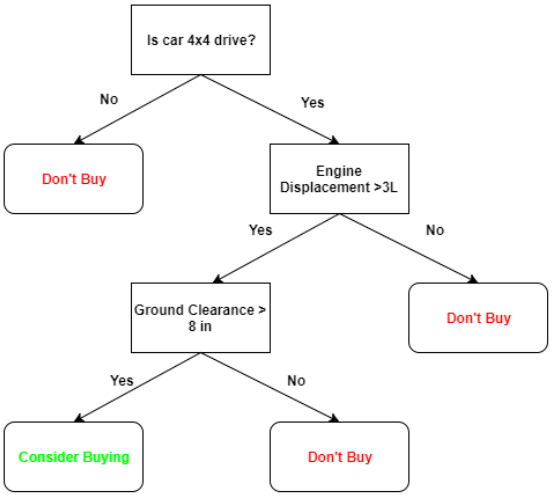


Fig. 1 Decision tree Classifier

B. Random Forest (RF)

The only thing prevents tree to become ideal tool for predictive learning is inaccuracy. This means trees work great while we use them with the data which used to create them, but they are not flexible enough. This is the reason, random forest is considered better then decision as

It is adaptable, simple to use calculation, works without hyper parameter tuning. It is easy to use and one of the most used algorithm. It is used as both relapse and order assignment.

Working of Random Forest:

Random Forest is a Supervised machine learning algorithm. Which means we have to use labelled dataset. Random Forest, as its name suggests, it created random set of decision tree. Decision trees are made on the bases of random attributes. Random sampling of training data points are used when building trees. After getting different trees with different attributes, all the trees are merged in the end. That leads in increase in stability of the model. As the decisions are made on the random attribute, precision of result increases.

"For each decision tree Scit-learn calculates nodes importance, using Gini Importance.

Mathematica formula for Gini importance:

$$ni_j = w_j\ c_j - w_{Left(j)}c_{left(j)} - w_{right(j)}c_{right(j)}$$

It is considering 2 child nodes.

Ni(j) = importance of node j

W(j)=Weighted number of samples reaching node j

C(j)= Impurity level of node j

Left(j)=child node from left split on node j

Right(j)=child node from right split on node j".

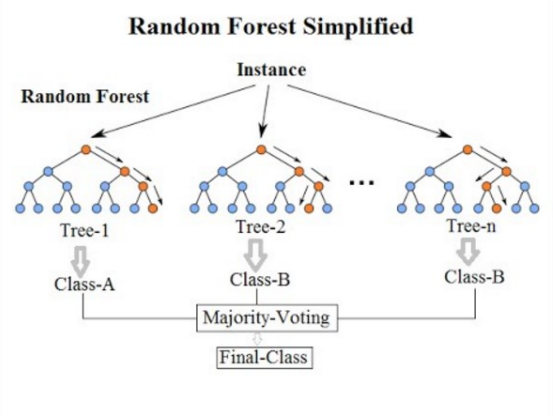


*Fig.2 Random Forest Classifier (*Wikimedia)

C. Support Vector Machine (SVM)

Support Vector machine Learning is a class of machine learning which works on supervised approach or on labelled data, in SVM the input data have already all the required result. SVM predicts good results in both classification and regression problems. Using SVM we can separate 2 classes points using a hyperplane. Center line between who distinct values of which we have classified is known as hyperplane. SVM has its own unique way of working as it just not only draw a hyperplane, it also generates two margin lines, which will be having some distance, so that it could be easily linearly separable for each of the classification points. Margin planes will be linear to hyperplane, and they will be passing through the nearest point we used to classify. Distance between margin plane and hyperplane is denoted using d+ and d-. Distance between two margin plane is known as Marginal distance. Using that Marginal distance, we can classify problem in much easier way. There colud be multiple hyperplane and marginal distance, but we are supposed to get the marginal distance with maximum margin, so that classification becomes easy. Support vectors are the are points from which the margin line passes by.

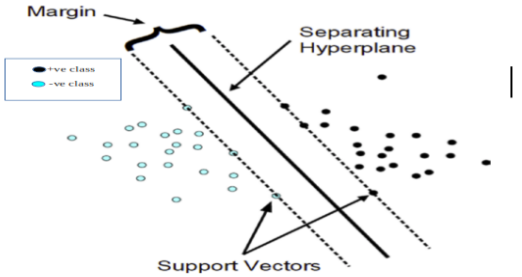


Fig.3 Support Vector Machine Classifier

$$h_{w,b}(x) = g(w^T x + b)$$

Class labels are denoted as -1 for negative class and +1 for positive class in SVM.

$$y \epsilon \{-1,1\}$$

The final optimization problem that we get fitting the best parameters:

$$min \frac{1}{2} ||w||^2$$
$$s.t. y_i(w.x_i + b) \geq 1, \forall x_i$$

D. Naïve Bayes (NB)

Naïve Bayes is a Supervised Machine learning algorithm, Input data type is labelled. Naïve Bayes algorithm is based on Bayes algorithm, or we can say that the Naïve Bayes algorithm is Advance version or Bayes algorithm. Naïve Bayes works on conditional probability, it means it a possibility is dependent on the probability on fulfilling of other events or the relationship with another event.
Working of Naïve Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Using this equation, we do find Naïve Bayes probability of occurring A. Here A is the resultant value and B is on which A is dependent. Using this we can find probability of A(P(A)), given B is true. B is termed as evidence. Probability of A is the priori of A, which means probability of A even before evidence is seen. Evidence is taken as an attribute value of an unknown instance. Probability of occurring Event A with respect to B(P(A/B)) is a posteriori probability of B, that means possibility of event after evidence is seen.

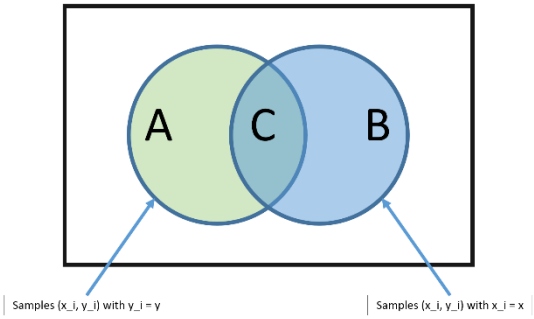


*Fig.4 Naïve Bayes Classifier (*cornell)

E. Logistic Regression (LR)

Logistic Regression is a Supervised machine learning approach, which means we can use this in both classification and regression. But Logistic Regression is popular for its use in classification [17]. It is used in Binary Classification. Logistic Regression has taken from one of the Statistics function, Logistic function also known as Sigmoid Function. Logistic Function is used to describe data in between 0 and 1, but never exactly at those limits. Logistic Function makes a S type curve as its resultant, which makes it very convenient function. Using that S type curve, it becomes easy to classify data.
It is used as categorical data when the variable is dependent. Like linear regression an equation is used. Input values are combined linearly using weights to predict an output value. Key difference between Linear and logistic regression is that, in logistic regression, we get output value in binary on the other hand in Linear regression we get output value as numeric.
In Logistic Regression, we use a cost function:

$$Cost(h_0(x), y) = \begin{cases} -\log(h_0(x)) & if\ y = 1 \\ -\log(1 - h_0(x)) & if y = 0 \end{cases}$$

Cost function is used so that we can develop model with high precision and minimise error. Average of loss function is considered as cost function. The loss function is a value which is calculated at every instance.

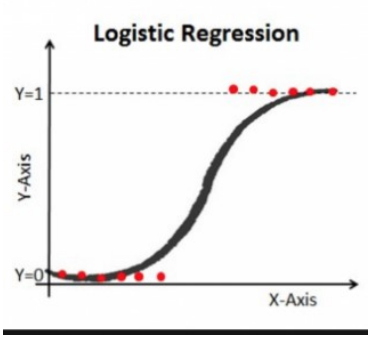


Fig.5 Logistic Regression Classifier (ODSC)

F. XGBoost

XGBoost is a gradient boosting framework-based algorithm. It works somewhat similar to decision tree.It can be used to predict various type of information, which may be unstructured, such as image, text. It is more efficient when it comes to small to medium data size, and even can outperform neural network. It was developed as research project at the University of Washington [18]. Strong area of this algorithm is ranking, regression and classification. It was developed to use resources such as time, computation power and memory efficiently. It have some different features too, which makes it one of the unique algorithm.
Sparse Aware: Using this it can automatically handle missing values.
Block Structure: Using this many trees can be generated simultaneously.
Continued training: Using this algorithm can further boost an already fitted model on new data.

G. Gradient Boosting Classifier

Gradient Boosting is a one of the most powerful technique to build predictive model. It is based on the Adabost or adaptive boosting. It is based on the concept of whether a weak learner, combined can become better. The weak learners in AdaBoost are decision trees with a single split, called decision stumps for their shortness.
Working of Gradient Boosting:
Gradient Boosting uses three steps:
Loss Function: Loss function is dependent on how we are using this algorithm. There are two types of loss function, which are used by Gradient Boosting Algorithm, squared error, and logarithmic loss. While performing regression, squared error is used and for classification logarithmic loss is used.

Weak Learner : Decision tree, precisely regression trees are used as weak learner. Using different trees we get output real values for splits and whose output can be added together, allowing which using subsequent models output can be added and using which we can make model more precise and near to the actual value.

Additive Model: Using this Gradient Boosting, add trees without touching the existing one. To minimise the loss while adding tree, gradient descent procedure is used. Gradient descent allows algorithm to minimize the set of parameters. These parameters include weights in neural network and confident in a regression equation. After the loss is calculated, while performing Gradient descent procedure, to reduce loss we add we must add a tree. This approach is known as Functional descent or gradient descent with function.

H. Artificial Neural Network (ANN)

Neural Network term is derived from the term Network of neuron. Base of this concept is Human brain, which have zillions of neurons, which can work simultaneously, perform different action, and capable of learning from outside using experiences. Artificial neural network is also supposed to work similarly, which have capacity to learn from experiences. Experience to the Artificial neural network is taught by providing dataset, using which ANN prepares some model after finding some common pattern between all the information provided. ANN is considered to one of the most efficient, reliable, and powerful approach towards machine learning. Accuracy of ANN is much more higher then the traditional approaches, because ANN tries to learn using different aspects towards data, which helps in predicting better results then other. Neurons in ANN is designed in similar way as of humans.

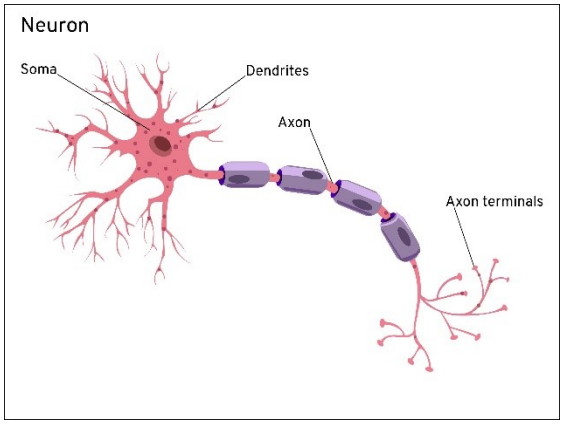


Fig.6 Human Neuron(lidolearning)

Human's neuron have dendrites which fetches data, then there is a cell body, which helps in processing the fetched data of dendrites, then there is Axom, using which we can pass our data to terminal axom or synapse, which is further connected to another dendrites. By this way whole Neural network in human is made. There are zillions of such neurons using which our brain works. In machine too, we try to make similar structure.

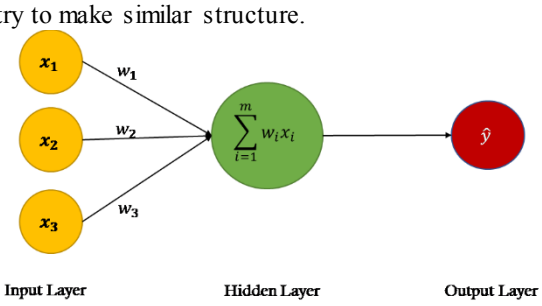


Fig.7 Artificial Neuron (Vaibhav sahu)

How Artificial Neural Network Works:
In machine similarly we provide input as a dataset to the neurons. Neural network is made up of Layers of Neurons. First layer is known as Input layer. Last layer is known as Output layer and in between layers are known as hidden layers. Hidden layers are responsible for most of the output we receive. Firstly, data is transmitted to the first layer of neurons. Neuron of first layer are connected to the next layer using channels [7]. Each channel is assigned some weight. Using weights, we decide the priority or the value of some input value. While processing First layer receives input and the input is multiplied by some weights (value of channel). And then the sum of all the product of input value and weights are sent to hidden layer which have some value known as bias. Then the value of bias is added to the input sum [16]. This value then passed to a threshold function known as activation function. That activation function decides if the neuron will be activated or not. If the neuron is activated, then it will transmit data to the next neuron using channels. This process is known as forward propagation. The output layer the neuron with highest value fires and determines the output.

These are just probability. This prediction may be right or wrong. Machine compares the given output with the actual output. Using this machine checks how much percent precited output is matched with the actual output.

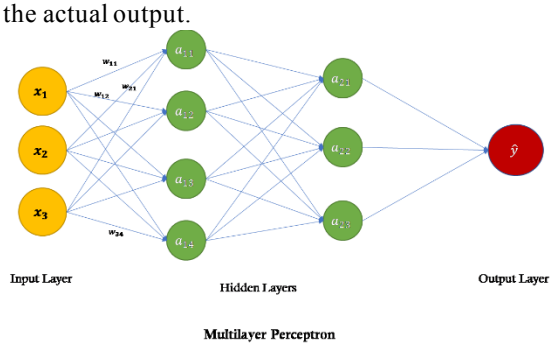


Fig.8 Working of Artificial Neuron(Vaibhav sahu)

This data then transmitted back to the network to reduce error. This is known as back propagation. Using these information weights are adjusted. This forward and backward propagation performs using various inputs. This happens until weights are adjusted correctly. This forward and backward propagation is defined by epochs.

DATA ANALYSIS

Data analysis is one of the crucial stages of Research. It is a somthing used by researchers to reduce information to a story which is easy to communicate and can be easily use to derive results and insights [15]. It is a process which requires statistical and logical techniques. After applying some techniques we are supposed to evaluate or analyse data, by which we can get some useful information, suggesting conclusion, and supporting decision making

There were two phase of experiment for this study: (1) training phase, and (2) test phase. We divided data in 70/30 Ratio. Which means we used 70 % of data in training and 30% in testing. The parameters considered in the experiments were as follows: (1) Sensitivity, (2) Specificity, (3) Precision, (4) Negative Predicted Value (NPV), (5) F1 Score, (6) False Negative Rate (FNR), (7) False Positive Rate (FPR), (8) False Discovery Rate (FDR), (9) False Omission Rate (FOR), (10) Accuracy

PERFORMANCE AND RESULT ANALYSIS

Description of different Parameter considered to evaluate the accuracy of model:

Sensitivity: Percentage of positive cases identified correctly.

Specificity: This measures proportion of correctly identified negatives

Precision: This is the percentage of correct prediction. It defines the ratio of true positive to the sum of true positive and false positive.

Negative Predicted Value (NPV): It is a value which is negative, and machine predicted the same.

F1 Score: Percentage of correct positive prediction.

False Negative Rate (FNR): It is a value, which is true but stated false by machine.

False Positive Rate (FPR): It is a value Which was initially false but stated true by machine.

False Discovery Rate (FDR): While conducting multiple comparison It is used to conceptualize type 1 errors in null hypothesis.

False Omission Rate (FOR): It is a proportion for which true value is positive and predicted result is negative.

Accuracy: It is a ratio of correctly predicted observation to the total observation.

Decision tree (DT)

Sensitivity	Specificity	Precision	NPV	F1 Score
0.8787	0.1724	0.8717	0.1818	0.8752
FNR	**FPR**	**FDR**	**FOR**	**Accuracy**
0.1213	0.8276	0.1283	0.8182	0.7832

Table 1: Evaluation Parameters for Decision tree

Random Forest (RF)

Sensitivity	Specificity	Precision	NPV	F1 Score
0.8735	0.5	0.9973	0.0182	0.9313
FNR	**FPR**	**FDR**	**FOR**	**Accuracy**
0.1265	0.5	0.0027	0.9818	0.8718

Table 2: Evaluation Parameters for Random Forest

Support Vector Machine (SVM)

Sensitivity	Specificity	Precision	NPV	F1 Score
0.8738	-	1	0	0.9327
FNR	**FPR**	**FDR**	**FOR**	**Accuracy**
0.1262	-	0	1	0.8738

Table 3: Evaluation Parameters for support Vector

Naïve Bayes (NB)

Sensitivity	Specificity	Precision	NPV	F1 Score
0.8654	0.1273	0.1203	0.8727	0.2113
FNR	**FPR**	**FDR**	**FOR**	**Accuracy**
0.1346	0.8727	0.8797	0.1273	0.2168

Table 4: Evaluation Parameters for Naïve Bayes

Logistic Regression (LR)

Sensitivity	Specificity	Precision	NPV	F1 Score
0.87	0	0.984	0	0.9235
FNR	**FPR**	**FDR**	**FOR**	**Accuracy**
0.13	1	0.016	1	0.8578

Table 5: Evaluation Parameters for Logistic Regression

XGBoost(XGB)

Sensitivity	Specificity	Precision	NPV	F1 Scor
0.8726	0.0769	0.9686	0.0182	0.9181
FNR	**FPR**	**FDR**	**FOR**	**Accurac**
0.1274	0.9231	0.0314	0.9818	0.849

Table 6: Evaluation Parameters for XGBoost

Gradient Boosting Classifier (GBC)

Sensitivity	Specificity	Precision	NPV	F1 Score
0.8702	0.0769	0.9679	0.0182	0.9165
FNR	**FPR**	**FDR**	**FOR**	**Accuracy**
0.1298	0.9231	0.0321	0.9818	0.8462

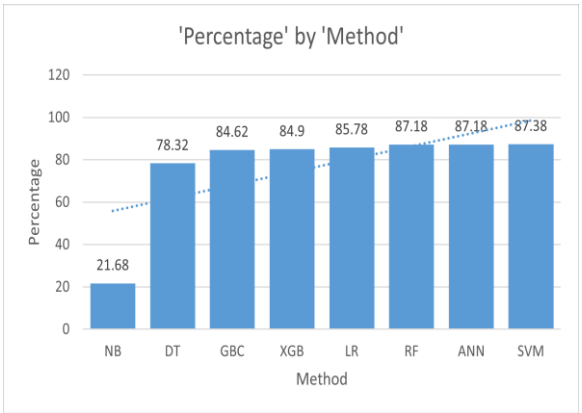
Table 7: Evaluation Parameters for Gradient Boost Classifier

Artificial Neural Network (ANN)

Sensitivity	Specificity	Precision	NPV	F1 Score
0.8718	-	1	0	0.9315
FNR	**FPR**	**FDR**	**FOR**	**Accuracy**
0.1282	-	0	1	0.8718

Table 8: Evaluation Parameters for Artificial Neural Network

Accuracy comparison of all the method applied



Bar Graph 1: Comparing results of all applied method

Table findings: We found that using various algorithms Sensitivity, which means Percentage of positive cases identified correctly was highest of Decision tree, i.e., 87.87%. Highest Specificity, which measures proportion of correctly identified negatives was of Random Forest i.e., 50%. Highest Precision, that is percentage of correct prediction. It defines the ratio of true positive to the sum of true positive and false positive was of both ANN as well SVM that is 100%. Highest Negative Predicted Value (NPV), that is a value which is negative, and machine predicted the same was highest of Naïve Byes, that is 87.27%. Highest F1 score was of Gradient boosting classifier, 91.81%. Lowest FNR was of Decision tree,12.13%. Lowest FPR was of Random Forest, that is 50%. Lowest FDR was of ANN and SVM, that is 0%. Lowest FOR was of Naïve Bayes, 12.73%. And best accuracy overall was given by SVM, that is 87.38%.

The dataset we used had 76 attributes. As we used eight algorithms to fetch best results, the best we found was 87.38% and worst was 21.67%. Best was using Support Vector Machine (SVM). The worst result we got was using Naïve Bayes (NB).

These algorithm and procedure can be further used to detect mental health of individual even before it gets too late. The best result we got was 87.38%

CONCLUSION

This paper shows a utilization of various Machine Learning calculations for determining the mental health of an individual, using various common know attribute, which are even available in numeric values. This attribute contains earning and spending pattern of individual, also household condition family members and investment was also an important part of the dataset. Which enabled us to make such kind of model, which can predict mental health on the bases of income and expenditure.

The examination of the outcomes connotes that the model can be used to get required information, which is about 87.38 percent correct. Further this research can be used in further improvement of the system or making it more efficient.

References

[1] Reddy, U. Srinivasulu, Aditya Vivek Thota, and A. Dharun. "Machine learning techniques for stress prediction in working employees." *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. IEEE, 2018.

[2] Shatte, Adrian BR, Delyse M. Hutchinson, and Samantha J. Teague. "Machine learning in mental health: a scoping review of methods and applications." *Psychological medicine* 49.9 (2019): 1426-1448.

[3] Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS medicine, 2(10), e267.

[4] Mohr, David C., Mi Zhang, and Stephen M. Schueller. "Personal sensing: understanding mental health using ubiquitous sensors and machine learning." *Annual review of clinical psychology* 13 (2017): 23-47.

[5] American Psychological Association. (2012). Gender and stress. https://www.apa.org/news/press/releases/stress/2010/gender-stress

[6] Garcia-Ceja, Enrique, et al. "Mental health monitoring with multimodal sensing and machine learning: A survey." *Pervasive and Mobile Computing* 51 (2018): 1-26

[7] Chancellor, Stevie, Eric PS Baumer, and Munmun De Choudhury. "Who is the" Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media." *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019): 1-32

[8] Graham, Sarah, et al. "Artificial intelligence for mental health and mental illnesses: an overview." *Current psychiatry reports* 21.11 (2019): 1-18.

[9] Bhattacharyya, R., & Basu, S. (2018). India Inc looks to deal with rising stress in employees. Retrieved from 'The Economic Times'.

[10] WHO. (2019, September 2). Suicide. https://www.who.int/news-room/fact-sheets/detail/suicide

[11] b_depression. (2018, November 17). [Dataset]. frankcc. https://www.kaggle.com/francispython/b-depression.

[12] Shwetha, S, Sahil, A, Anant Kumar J, (2017) Predictive analysis using classification techniques in healthcare domain, International Journal of Linguistics & Computing Research, ISSN: 2456-8848, Vol. I, Issue. I, June-2017

[13] Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266.

[14] Pang Z, Yuan H, Zhang Y-T, Packirisamy M. Guest Editorial Health Engineering Driven by the Industry 4.0 for Aging Society. IEEE J Biomed Heal Informatics. 2018;22(6):1709–10.

[15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.

[16] Bone, Daniel, et al. "Signal processing and machine learning for mental health research and clinical applications [perspectives]." *IEEE Signal Processing Magazine* 34.5 (2017): 196-195.

[17] T. Jain, V. K. Verma, M. Agarwal, A. Yadav and A. Jain, "Supervised Machine Learning Approach For The Prediction of Breast Cancer," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-6, doi: 10.1109/ICSCAN49426.2020.9262403.

[18] Jain, Tarun, et al. "Performance Prediction for Crop Irrigation Using Different Machine Learning Approaches." Examining the Impact of Deep Learning and IoT on Multi-Industry Applications. IGI Global, 2021. 61-79.