A Data Analytics Approach to Screen Depression using Beck Depression Inventory – II

[1]Priyanka Priyadarshni Pradhan, [2]Sunil Dhimal and [3]Kalpana Sharma

[1,2,3]Sikkim Manipal Institute of Technology, Majitar, Rangpo Sikkim – 737136
Email: priyankapradhan930@gmail.com

Abstract: Depression is one of the common disorders prevalent in current population. Depression is described as a mental health disorder with consistent depressed mood, lack of interest in day to day activities which intern paralyzes normal daily life. Technical intervention to screen depression in non-clinical population which records, classify depression among the population and provide indicators which helps in improving the diagnosis process is the main area of the study. This paper discusses the use of Beck Depression Inventory – II (BDI-II) for screening a population with depression along with its severity. BDI-II is 21 questions questionnaire that has been used to screen depression. These questions which are the variables are scored to screen depression. Best suited classification algorithm is compared on the basis of optimal sensitivity and specificity and good accuracy on the provided BDI response data.

Index Terms: Accuracy, classification techniques, depression, specificity, sensitivity

I. INTRODUCTION

Depression screening can be used to improve diagnosis of depression as screening comprises of identifying indicators. There isn't any prevalent lab test to screen depression in Sikkim. Clinical data for confirming depression is merely available in the form of digital data and difficult to acquire. There are a number of questionnaires that are used for screening depression. Beck Depression Inventory - II (BDI - II) is one of the questionnaires that are used for screening subjects with depression [1].

BDI - II is used as a questionnaire to conduct the analysis as the questionnaire fits well with age group of 13 and above. BDI - II is a world-wide used self-rating scale for measuring depression [2]. BDI has been used in non-clinical subjects [3]. BDI - II is also used in Sikkim and is used for initial screening of depression among the patients.

Analysis using machine learning has been an area of research for diagnosis of mental disorders which diagnose a disorder with considerable level of accuracy. Studies have been made for improving the classification of depression by adding dimensionality to the classification than by just having pure categorical classification [12]. There are researches conducted for classification of disorders like Major Depressive Disorder, Bipolar Disorder and Schizophrenia and Normal subjects using the data collected from electroencephalogram signals [11], classification of the disorder is based on maximum likelihood principle. Machine learning algorithms have been used for classification of depressive disorders using ensemble techniques [10] which comprises of algorithms like multi-layer perceptron, KNN and SVM and the accuracy was measured using confusion matrix.

Detection of depression from a given set of question involves those features that have contributed the most for detection of depression, these features needs to be identified which discriminates the class of depression. There are on-going researches which may help in improving the classification of depression for screening of depression. Quality indicators help in improving the diagnosis process for depression.

II. RELATED WORKS

There are number of work done and research are conducted in accordance to diagnosis of depressive disorders. Ahmad Khodayari-Rostamabad, James P. Reilly, GareyHasey, Hubert deBruin and Duncan MacCrimmon[11] proposed an automated diagnosis procedure for diagnosing psychiatric disorder based on statistical machine learning using electroencephalogram (EEG). The features obtained from the EEG signals were statistical quantities which were reduced to important features that were used to evaluate the class of psychiatric disorder. The evaluation was done using maximum likelihood principle for four classes of psychiatric disorders namely Major Depressive Disorder, Schizophrenia, Bipolar Disorder and Normal subjects. The probabilistic event that the reduced feature belongs to a particular class was based on maximum likelihood classification rule. Subhrangsu Mukherjee, Kumar Ashish, NirmalBaranHui, Subhagata Chattopadhyay [13] discussed modeling of depressive datausing intelligent neural network for accurate diagnosis of depression, two approaches namely Back Propagation Neural Network (BPNN) and Radial Basis Neural Network (RBNN) are discussed for modelingmedical depression data.Three fully connected feed forward neural network model was used. The degree of depression was given in the output layer which had one neuron.BPNN and RBNN used different activation function for modelling depression data and finally gave the degree or severity of depression as an output. Log sigmoid and radial basis function was used as a transfer function in the hidden layer for BPNN and RBNN respectively. The output layer used linear transfer function for RBNN and tan sigmoid function for BPNN. RBNN gave better results for prediction compared to the standard BPNN for the given set of data however the data was not enough to infer such results. The model ling was done using the clinical data. B. Ojeme, M. Akazue and E. Nwelih[14] used ensemble techniques which were a combination of machine learning algorithm like K-nearest neighbour, Support Vector Machine, Artificial Neural Network as an ensemble technique for classification of depressive disorder. The result of the machine learning technique, trained independently and jointly. The ensemble method, in different combination show minor but consistent improvement. The result calculated using confusion matrix. The strength of ensemble technique to automatically detect depression was better. The data used for the analysis were clinical data, features were extracted which involved various social and economic aspects of the patients which were characterized as symptoms. The classifier was tested individually and jointly to check for the accuracy. The main gap for this paper is that comorbidity issue is not handled with the given set of data. JiayueCai, Z. Jane Wang and SilkeAppel-Cresswell, Martin J. Mckeown[15] discussed the use of Beck Depression Inventory-II to screen depression in patients suffering from Parkinson'disease (PD) using an appropriate feature selection algorithm.Jiayue investigates subset of BDI-II items for identification of depression. 21 questions responses were analyzed for subjects suffering from PD, relevant features or subset of features were selected using the filter method for selecting the features which ranked the features. The feature selection method used minimal-redundancy-maximal-relevance (mRMR) which focused on the relevance of the feature with the class it did not consider the internal dependency among the features if it existed which minimized the redundancyand maximize the dependency of the features with the target class label. To examine performance of selected features, Support Vector Machine (SVM) was used. Studies are conducted with clinical data and EEG signals to screen depression and also responses of the questionnaires are used to screen depression. In non-clinical population depression screening can be done using questionnaires.

III. DESIGN STRATEGY

Lifecycle model for healthcare analytics comprises of the discovery phase, where the problem is identified, it learns about the domain including relevant history of projects in the specified areas. The second phase is the data preparation phase which needs an answer the following question, whether quality of data is good to be used for subsequent processes. Model planning phase determines the methods and techniques that will be used. Model building phase team develops datasets for testing and training for production purpose. The result needs to be communicated. Domain knowledge is important to tackle the problem using technical help. Data collection and preparing data, data needs to be made uniform as the data reflects the solution to the problem, Support Vector Machine model is used for classification as the performance of SVM is better for BDI-II response data. Classification process is used to classify the features into labels [4]. Data set are divided into two sets training and validating sets to measure the performance of the selected features and factors. Performance of the predictive power is measured or evaluated using confusion matrix and is measured in terms of accuracy, specificity and sensitivity [4].

IV. SOLUTION STRATEGY

A. Classification of BDI-II Responses

Classification of BDI-II responses using a supervised machine learning technique classifies data with good accuracy [10]. The selection of model is based on which algorithm performs better in a given situation given below are few classifiers used for classification of depression data. Classification algorithms were chosen considering two important factors that is the interpretability andaccuracy. Accuracy, specificity and sensitivity of the data are the performance measures used to evaluate the model for classification. The classification method that has been used for classifying BDI-II responses on the basis of depressive severity are Naïve Bayes Classifier, Decision Tree, Support Vector Machine and K-Nearest Neighbor.

Naïve Bayes Classifier is one of the simplest and an efficient method for machine learning classification, primarily used for text classification or classifying new article. Naïve Bayes Classifier has an assumption that all the features that are considered are independent to each other, it has a probabilistic approach for classification It is based on Bayesian theorem [2] given by

$$P(C|X) = \frac{P(X|C)*P(C)}{P(X)},$$

with an assumption that the variables or predictors are independent to each other. It calculates the posterior probability of an occurrence of a class when a given set of features are given. Classification is done on the basis of which class has the highest probability and is known as Maximum A Posteriori (MAP). Naïve Bayes classifier works under a very strong assumption that the features are independent to each other, so if there exist a relationship between the features, Naïve Bayes classifier takes the features as an independent feature without considering its dependency with the other features and performs the classification. However for classification purpose is the only objective, then this assumption may not be an issue. Disadvantage of Naïve Bayes classifier is when the dataset is small the precision of the algorithm may decrease. Another disadvantage is that if the feature value for a class is not there then the likelihood of the feature for that class is 0 which affects the posterior probabilityand prediction may not be correct.

Another powerful classification algorithm is Support Vector Machine (SVM). It is basically used for binary classification problem where it has two classes and these two classes are used to segregate using a hyper plane

[10]. The hyper plane separates the classes distinctively for more than two classes. A hyper plane should be such that it separates the classes and also the distance between the hyper plane and the support vectors should be more. SVM has a kernel trick which helps in separation of classes where it takes low dimensional input space and transforms into a high dimensional space. It converts non separable problem to separable problem by using function called kernels. It is more suitable in non-linearly separable problem. SVM is inherently a two class classifier. SVM can be extended for multiclass classification by using One Versus All (OVA) technique or One Versus One (OVO). Since the BDI-II responses were classified using SVM which gave good accuracy with balanced sensitivity and specificity.

KNN stands for k nearest neighbor, it an algorithm which does not have strong assumptions like Naïve Bayes. Due to its non-parametric property KNN has been used for solving real world classification problems unlike other algorithms which have assumptions. KNN classifier finds it's k nearest neighbor where k is number of neighbor data points. Each point is given weights which are calculated using the distance and is known as weighted KNN. Weight of the point shares an inverse relation with distance of the point to be classified. Near points have higher weight than the farther points [10].

Decision tree is a non-linear classifier; it visualize the relationship among the features and its outcome by maintaining a tree structure. The decision tree classifies data in a dataset starting from the root it flows to the leaves which represents one class [10]. Task is to predict or classify a new set of data based on the decisions made from the attributes.Decision trees helps in prediction in a very different manner as it also tries to understand the features or the attributes on the basis of which decision is made. Decision tree algorithm split the data into subsets on the basis of the attributes of the data. Then it checks if subsets are pure or not (all belonging to same class) if yes then the algorithm stops if not then sub setting process is repeated.

Which attribute to split on – While splitting purity of the subset is checked. Ideally attribute with all pure subset is selected. Metric that will measure for the purity of subset and one way is measure of uncertainty. Entropy is the measure of the uncertainty of a class in a subset of example. Given by $H(S) = -p(+) \log2 p(+)-p(+) \log2 p(+)$ bits where $p(+)$ is the probability of positive in the subset.

Decision tree is interpretable; it has a concise description of what makes an item to be classified in a particular class. The advantage of Decision tree is that it has no underlying assumption for the model.

V. IMPLEMENTATION

The data was collected using a Google form which captured the responses from non-clinical subjects, the link to the Google form for Beck Depression Inventory – II (BDI -II) is http://bit.ly/2yYvq02.The responses were then saved as an excel sheet, which comprises of responses of the BDI-II questionnaire. The responses were scored as per the guidelines for each question and the total score was calculated and then assigned a class of severity of depression.Supervised classification algorithms like SVM, KNN, Decision Tree and Naïve Bayes were used to classify the collected data set. R software is used to perform the classification of BDI responses [16].

405responses was collected and divided into training and testing data in the ratio 7:3 respectively.

Cross validation with 10 folds was performed on the training data set that was acquired after splitting the original data set.A model of the training data set was obtained after the cross validation process using the chosen algorithm.This training model was set as the reference for prediction for the testing data set.To measure the performance of the algorithm confusion matrix was created. The best classification algorithm amongst the four algorithms stated above was chosen on the basis of optimal sensitivity and specificity and good accuracy.

VI. ALGORITHM COMPARISON

Table 1, gives the comparison of classification algorithm used for classification of BDI-II responses with its advantage and disadvantages and its relevance with the classification of BDI-II responses

Table 1.COMPARISON OF CLASSIFICATION TECHNIQUES FOR BDI-II RESPONSES

Algorithm	Description	Relevance with BDI-II response data
Naïve Bayes	Simple and efficient, features are independent	Works well if classification is the only issue. Accuracy is as high as 80%
SVM	Strong classifier with kernel function. Relevant for real world problems	Accuracy is high as 90% using OVO
KNN	No assumption, inherently a multiclass classifier. Problem in determining the value of k. Uncertainty in calculating the distance	Noassumptions, accuracy is high as 71%
Decision	Determines discriminant	Identification of discriminating features.

Tree	features. Handles both discrete and continuous data	Accuracy is 56%

VII. PERFORMANCE OF CLASSIFICATION TECHNIQUES FOR BDI-II RESPONSES

With the given data sets classification algorithm that is appropriate for BDI-II responses is the Support Vector Machine with an accuracy of 0.943 and a balanced specificity and sensitivity, in case of Beck Depression Inventory false negatives needs to be reduced as much as possible. The performance of the algorithm computed using a confusion matrix which measures the performance using accuracy, specificity and sensitivity as performance metrics is given in Table 2.

Table2.CLASSIFICATION TECHNIQUES PERFORMANCES FOR BDI-II RESPONSES

Algorithm	Accuracy	Specificity	Sensitivity
Naïve Bayes	0.836	Minimal = 0.896 Mild = 1.000 Moderate = 0.500 Severe = 0.710	Minimal = 1.000 Mild = 0.857 Moderate = 0.953 Severe = 1.000
SVM	0.943	Minimal = 0.800 Mild = 1.000 Moderate = 0.800 Severe = 1.000	Minimal = 1.000 Mild = 0.950 Moderate = 0.979 Severe = 0.978
KNN	0.709	Minimal = 0.788 Mild = 0.250 Moderate = 0.000 Severe = 1.000	Minimal = 1.000 Mild = 0.814 Moderate = 0.901 Severe = 0.967
Decision Tree	0.566	Minimal = 0.280 Mild = 0.820 Moderate = 0.200 Severe = 0.000	Minimal = 0.826 Mild = 0.692 Moderate = 1.000 Severe = 0.820

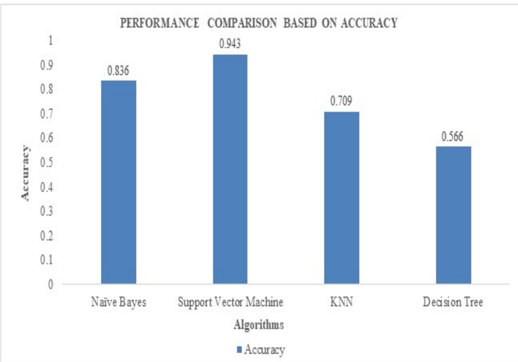

Fig. 1. Algorithm Comparison based on Accuracy

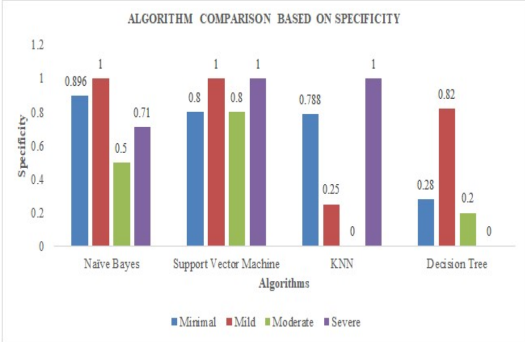

Fig. 2. Algorithm Comparison based on Specificity

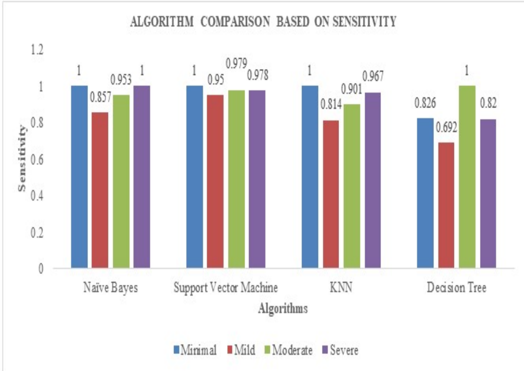

Fig. 3. Algorithm Comparison based on Sensitivity

VIII. CONCLUSION

Evaluating and interpreting BDI-II responses is done using a sample of non-clinical population. The respondents age range from 17 to 30. Data was collected using a Google form. The collected data was successfully classified using four classification algorithms viz. K Nearest Neighbor, Support Vector Machine, Naïve Bayes and Decision tree. On the basis of the observed results it could be concluded that Support Vector Machine is the best suited classification algorithm amongst the aforementioned algorithms for the collected BDI – II response data set. Support Vector Machine is a strong classification algorithm that has been used to solve classification problem in a diverse domains of application.

IX. FUTURE SCOPE

For enhancing the results obtained, the data set can further be processed using feature extraction and selection process where, the features that contributes in classifying depression is identified and selected using the feature selection method which ranks the features according to its importance on how well it predicts depression severity. To summarize the underlying pattern of the responses of the respondents, factor analysis (also known as feature extraction), an interdependence techniques, is used to evaluate the interdependence among the features or variables and create factors that represent the underlying pattern of the given data set which can be used for decision making process.

REFERENCES

[1] Understanding Depression, National Institute of Mental Health, “https://www.nimh.nih.gov/health/topics/depression/index.shtml”

[2] Working of Naïve Bayes Classifier Machine Learning “http://dataaspirant.com/2017/02/06/naive-bayes-classifier-machine-learning/”

[3] Parikhshit Sondhi “Feature Construction Methods: A Survey”

[4] Sagar S. Nikam “A Comparitive Study of Classification Techniques in Data Mining Algorithm” International Research Journal of Computer Science and Technology

[5] E. Stengel, M.D., F.R.C.P. Classification of Mental Disorders, Bull. Org. mond. Sante, Bull. WldHlth Org., 1959, 21, 601-663

[6] P. Richter, J. Werner, A. Heerlein, A. Kraus and H. Sauer, “On the validity of the Beck Depression Inventory,” Psychopathology, vol.31, pp.160–168, 1998

[7] Fabrigar, LeandreR.,Wegener, Duane T., Mac Callum, Robert C.,Strahan, Erin J. “Evaluating the use of exploratory factor analysis in psychological research” Psychological Methods, Vol 4(3), Sep 1999, 272-299

[8] Anna B. Costello and Jason W. Osborne’s “Best Practises in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis” Volume 10 Number 7, July 2005

[9] Timothy A. Brown and David H. Barlow.: Dimensional Versus Categorical Classification of Mental Disorders in the Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders and Beyond: Comment on Special, Journal of Abnormal Psychology 2005, Vol. 114, No. 4, 551-556

[10] S. B. Kotsiantis.: Supervised Machine Learning: A Review of Classification Techniques Informatica 31 (2007) 249-268 249

[11] Ahmad Khodayari-Rostamabad, James P. Reilly, GareyHasey, Hubert deBruin and Duncan MacCrimmon.: Diagnosis of Psychiatric Disorders Using EEG Data and Employing a Statistical Decision Model, 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argetina, August 31 – September 4, 2010

[12] D. Ignjatovic-Ristic; D Hinic, J Jovic, “Evaluation of Beck Depression Inventory in a nonclinical student sample” West Indian Medical Journal. Vol. 61, no.5 2012

[13] Subhrangsu Mukherjee, Kumar Ashish, Nirmal BaranHui, Subhagata Chattopadhyay.: Modelling Depression Data: Feed Forward Neural Network vs. Radial Basis Function Neural Network, American Journal of Biomedical Science, 2014, 6(3), 166-174; doi: 10.5099/aj140300166

[14] B. Ojeme, M. Akazue and E. Nwelih..: Automatic Diagnosis of Depressive Disorders using Ensemble Techniques Vol 8. No. 3(2), October 2015, 2006-1781

[15] JiayueCai, Z. Jane Wang and SilkeAppel-Cresswell, Martin J. Mckeown.: Feature Selection to Simplify BDI for Efficient Depression Identififcation, 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) 978-1-4673-8721-7/16

[16] https://www.r-project.org