Journal of Medical Systems (2018) 42: 88
https://doi.org/10.1007/s10916-018-0934-5

EDUCATION & TRAINING

Behavioral Modeling for Mental Health using Machine Learning Algorithms

M. Srividya[1] · S. Mohanavalli[1] · N. Bhalaji[1]

Received: 24 July 2017 / Accepted: 11 March 2018 / Published online: 3 April 2018

Abstract

Mental health is an indicator of emotional, psychological and social well-being of an individual. It determines how an individual thinks, feels and handle situations. Positive mental health helps one to work productively and realize their full potential. Mental health is important at every stage of life, from childhood and adolescence through adulthood. Many factors contribute to mental health problems which lead to mental illness like stress, social anxiety, depression, obsessive compulsive disorder, drug addiction, and personality disorders. It is becoming increasingly important to determine the onset of the mental illness to maintain proper life balance. The nature of machine learning algorithms and Artificial Intelligence (AI) can be fully harnessed for predicting the onset of mental illness. Such applications when implemented in real time will benefit the society by serving as a monitoring tool for individuals with deviant behavior. This research work proposes to apply various machine learning algorithms such as support vector machines, decision trees, naïve bayes classifier, K-nearest neighbor classifier and logistic regression to identify state of mental health in a target group. The responses obtained from the target group for the designed questionnaire were first subject to unsupervised learning techniques. The labels obtained as a result of clustering were validated by computing the Mean Opinion Score. These cluster labels were then used to build classifiers to predict the mental health of an individual. Population from various groups like high school students, college students and working professionals were considered as target groups. The research presents an analysis of applying the aforementioned machine learning algorithms on the target groups and also suggests directions for future work.

Keywords Mental health · Classification · Predictive analytics · Behavioral healthcare

Introduction

Mental well being of a person is the state of mind of that individual and also gives an overview of his/her general nature. Mental illness is a result of imbalances in brain chemistry. The assessment of mental wellness is also very critical to understanding and suggesting treatments to be given for patients with deviated mental behavior. The mental health of an individual serves as an indicator for effectively treating the ailments of the individual. It is essential to maintain the mental health profiles of different communities in order to predict any health related anomalies. The community can be broadly classified as high school adolescents, college goers and working professionals. There is a common notion that all categories of the population are commonly subject to stress and depression. It is a necessity to address the mental wellness of different categories at different times in order to prevent any serious illness. A executive board of World Health Organization (WHO) estimate in 2011 has predicted that by 2030, depression will be the leading cause of global disease burden. The radical shift to include the mental health profile of a patient by healthcare providers will be made mandatory in the coming years to provide better medication and also assist in faster recovery. Winters-Miner et al. [1] has discussed about the way medical predictive analytics will revolutionize the healthcare field globally.

This article is part of the Topical Collection on *Education & Training*

✉ S. Mohanavalli
mohanas@ssn.edu.in

M. Srividya
srividhyav@ssn.edu.in

N. Bhalaji
bhalajin@ssn.edu.in

1 SSN College of Engineering, Chennai, India

Psychologists focus extensively on psychotherapy and treating emotional and mental suffering in patients with behavioral intervention. Psychologists are also qualified to conduct psychological testing, which is critical in assessing a person's mental state and determining the most effective course of treatment. The developed prediction system will assist the psychologists in conducting psychological testing and in predicting the mental health of an individual. The psychologist and psychiatrist work in tandem to treat patient symptoms from both a behavioral and clinical standpoint. The fields of psychology and psychiatry are both essential in researching and developing treatment for improving mental and emotional health. According to WHO, 50 million Indians are suffering from depression which is one of the prevalent outcomes of mental illness. India has a total of only 898 clinical psychologists, one for every 1.3 million people and a total of 3800 psychiatrists, one for every 3,30,000. At this juncture it is very essential to provide mental health services available to larger community of population. Several factors that affect the general mental wellbeing of an individual include globalization, workplace pressure, competition at study place, etc. The proposed system is expected to perform behavioral profiling of individuals in an attempt to make mental healthcare more accessible.

With the advancement in technology the role of a psychologist can be supplemented or even replaced with Artificial Intelligence(AI) based mental health monitoring tools ranging from smart phone applications to wearable devices. The book by Poulin et al. [2] discusses the role of AI in behavorial and mental health care. It elaborates the use of AI for decision making and its applications for assessment and treatment. Machine learning is a type of AI which makes the computer to learn by training using large amounts of data without the need to explicitly program them. These algorithms parse data, learn from it and then make a determination or prediction about something. The machine learning algorithms play a very vital role in analyzing the data collected from the mentioned devices or tools. There are wide range of functionalities provided by these algorithms makes them suitable candidates for use in predicting mental health. The emerging field of 'predictive analytics in mental health' is becoming a reality with many moving towards using machine learning for aiding in preliminary clinical decision making process. The role of predictive analytics in mental health is discussed in Hahn et al. [3]. They have reported the applications, guidelines, challenges and perspectives of predictive analytics in this domain. There are a number of factors that influence the choice of the classifier used to build the predictive model. The classifiers like support vector machines, K-nearest neighbour, naive bayes, logistic regression could be used to build the model. The use of a ensemble of classifiers is also encouraged if the results obtained from individual classifier is not very satisfactory. This research work proposes to identify the mental health of an individual in a population using the classification algorithms. The primary objective of building a prediction model for assessing the menatl health is achieved by applying clustering and classification algorithms as discussed in the following sections.

Related work

The relationship of mental with the overall wellbeing of an individual has been under investigation by many countries for many years. Since the mental health is directly related to social and cultural aspects of different regions, it is very essential to custom make the decision support and prediction systems. In Bijl et al. [4] has studied the prevalence of psychiatric disorders in the Dutch population.

The mental wellness of an individual is directly related to early diagnosis of deviation in mental health. The WHO reports [5] the region wise status of different barriers in diagnosing mental health problems. The report urges the nations worldwide to harness the power of knowledge and technology to address the issue of mental health. Individuals suffering from abnormal behavior had to initially consult a psychologist for getting themselves diagnosed with the type mental illness they suffer from. The diagnosis was based on one to one interaction and followed by counseling sessions. Now with the advancement of technology there are various modalities to predict the state of mental health. WHO also discusses about the poor diagnosis and treatments provided globally in [6]. Though the state of mental health is a behavioral problem, it can lead to many illnesses if not diagnosed at an early stage.

Stress is defined as a state of psychological and physiological imbalance resulting from the disparity between situational demand and the individual's ability and motivation to meet those needs. It is a state of mental or emotional strain or tension resulting from an adverse or demanding circumstance. This leads to depression and anxiety. In L'Abate et al. [7] performed a detailed study about various mental illness, its causes and impacts, also suggesting the modalities that can be used for prediction and control.

Survey questionnaires [8], wearable sensors [9], bio signals are some of the modalities used for assessing mental health. We focus on the role of machine learning algorithms in assessing various aspects of mental health. Schaefer, Jonathan D., et al. in [10] studied the statistical relationship between mental health and other parameters like education attainment, socioeconomic attainment, life satisfaction, relationship quality. Various researchers have extensively studied social and behavorial characteristics on different types of individuals. Different categories of peoples are subject to disturbances in mental health due to different reasons. The factors that cause mental illness vary between different age groups of people. Some of the factors that contributes to mental illness for school students are the pressures of exams, Peer and parental pressure about marks scored and career options For working professionals the factors are work deadlines, promotion anxiety, monotony of job, financial position, family responsibility, etc.

The emergence of Internet of Things (IoT) helps us better asses individual and detect changes in behaviour by observing the shift in device usage pattern during different mental states. The work by [11–13] tells us the methods to handle such a scenario in a more efficient manner. It becomes increasingly important to monitor people with acute mental health problems. Such people will have to be carefully monitored and important alerts have to reach the health care providers to enable to act immediately in such situations. This security and risk assessment in health care sensor networks have been discussed in [14–16].

In Chinaveh et al.. [17] examined the effects of the problem-solving to enhance effective coping skills and psychological adjustment among Iranian college students. Hajiyakhchalireza et al.. [18] reported the effects of the problem-solving to enhance effective coping skills and psychological adjustment among Iranian college students. Aghaei et al [19] aimed to predict general health (GH) based on life orientation variables, quality of life (QoL), life satisfaction and age using questionnaires. Descriptive and multivariate regression analysis were used to analyze the data. In Strauss et al. [20] investigated the use of machine learning algorithms like cluster analysis, K-nearest neighbours (KNN), decision trees and support vector machines (SVM) for clinical forms analysis of mental health. Bio- index for tension, normal and relaxed state of individuals was done using bio sensors and classified using fuzzy logic and SVM in [21]. The use of kernel classification to build prediction models for image retrieval has been discussed by Wang et al.. in [22]. The use of ontology approach in building a multi-label prediction model for classification of anatomical therapeutic chemicals is suggested by Xiang et al. [23]. This paper also discusses the performance measures between the ontology based classifier and conventional classifier.

Rakesh et al [24] and Rebiero et al. [25] discuss the use of machine learning in suicide prediction. Kessler et al [26] reported the use of machine learning model built from survey predicted major depressive disorder (MDD) predicted its persistence with good accuracy. The work by Fleury et al [27] discusses the use of support vector machines (SVM) by quantifying the activities of daily living (AODL) by subjects that were considered. Smart health homes with various sensors installed were considered for this purpose.

The physiological signals like Electrocardiogram (ECG), Galvanic Skin Response (GSR), respiration, temperature, etc., have been used to detect mental stress. Smets et al. [28] investigates the use of six machine learning techniques for the detection of mental illness. It also reported that Bayesian networks and SVM gave reasonable accuracy. Mean opinion score (MOS) calculated for gauging the Quality of Experience (QOE). Generally QoE is used as a means to know the satisfaction of Quality of Service provided to consumers. In Xu et al. [29] studied the dependency between MOS and QoE. The higher the MOS, higher is the user's satisfaction/experience. QoE is used in this research to ascertain whether the identified labels are indeed correct. As suggested by Jung et al [30], K-Means is used to identify clusters from the dataset whose class labels are unknown.

Background study

The aim of this research work is to identify the individuals who are mentally distressed in the target population. These individuals need special attention in order to mentally balanced population. The basic form of getting to know about individuals in a population is to get responses for common benchmarked questions and rate them depending on their responses. This research is a pilot work for further in depth study about different kinds of mental illness. The target population 1 was decided to be high school/college going students and the target population 2 was working professionals with less than 5 years experience from different organisations. The questionnaire for the target group was framed in consultation with a psychologist. As per the expert's suggestion we decided to prepare the questionnaire to achieve the objective of identifying deviations in mental health. The final set of questions prepared was based on the work by Kern, et al. [31]. It supports five different factors that together support higher levels of well-being: engagement, perseverance, optimism, connectedness, and happiness. Based on scores obtained by respondents, a person's mental health status is predicted as mentally distressed, barely satisfied with life and optimistic.

Clustering

It is the process of organising similar objects into meaningful/useful groups such that the objects within a group are more similar than those in other groups. Clusters help understand data and to identify potential classes for further processing. Unsupervised learning is sometimes called clustering. It is a learning framework using a specific object functions, for example a function that minimizes the distances inside a cluster to keep the cluster tight. Cluster analysis has played a very vital role in psychology, pattern recognition, information retrieval, machine learning and datamining. An illness is a condition with a number of variations. Clustering is used to identify different types of illness. The different types of clustering techniques available are partitional, hierarchical, density based, grid based and model based methods. The clustering are discussed in detail by Glenn Milligan et al. [32].

Clustering is data exploratory data analysis tool that simply sorts data into groups without any explanation or interpretation. It basically is used for knowledge discovery and not prediction. The result of clustering is used to build taxonomies / class labels. The resulting numbers of clusters help in the task of assigning appropriate class labels. The resultant number of clusters are just representations of the possible class labels but cannot describe the mapping between the class label and the representative clusters. In short, clustering can be used a preprocessing step to determine the naturally occurring groups which can further be used build classifiers. It is very essential to validate the number of clusters obtained through simulation. The choice of appropriate clustering algorithm and related parameters depend on the individual data set and intended use of the results. A number of trials is required before deciding on the number of clusters. It is necessary to modify data preprocessing and model parameters until the result achieves the desired properties. There are a wide variety of cluster validity metrics used to ascertain the number of clusters obtained is indeed correct. The data point has to have high intraclass similarity and low intraclass similarity. High cohesion is expected between data points within the cluster and loose coupling is expected between adjacent clusters. This implies that the intraclass similarity has to be high and the inter class similarity has to be low. An exposure to clustering algorithms help is essential to know which can be used for the dataset under consideration. Extensive experiments on the dataset will pave way for choosing the appropriate clustering algorithm (s). Every algorithm follows a different set of rules for defining the similarity among data points.

The most commonly used clustering algorithms are: K-Means clustering, Hierarchical clustering, Density based clustering, K-medoids and its variants. K-Means clustering algorithm is an iterative algorithm which performs well under many real-world use cases and is one of the widely used algorithms. Let 'n' be the number of data points, 'K' be the number of clusters. Each of the 'n' data points are assigned to one of the 'K'. The main goal is to minimize the differences between data points within each cluster and maximize the differences between the clusters. Instead, the algorithm uses a heuristic process that finds locally optimal solutions. Put simply, this means that it starts with an initial guess for the cluster assignments, and then modifies the assignments slightly to see whether the changes improve the homogeneity within the clusters. This centroid based clustering assign the objects to the nearest cluster center, such that the squared distances from the cluster are minimized. To find the number of clusters in the data, the user needs to run the *K*-means clustering algorithm for a range of *K* values and compare the results.

Hierarchical Clustering is one of the predominantly used algorithms that aims to build hierarchy of clusters. A cluster tree is built to represent data where each group is linked to two or more successor groups. These groups are then nested and organised to form a tree. Here the total number of clusters is not predetermined before the start of tree creation. The two approaches used are: bottom up (agglomerative clustering) and top down (divisive clustering). The agglomerative approach merges clusters iteratively whereas the divisive approach does the reverse by starting with all objects in one cluster and subdividing them into smaller pieces. If the merge or split decision is not chosen well then it results in low quality clusters.

Density Based Clustering is used to identify non spherical clusters and outliers in data. The clusters are regarded a dense region of data points which are separated by regions of low density. The number of data points and the radius from the chosen data point determine the efficiency of the algorithm. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is most widely used density based algorithm.

K- Medoids clustering or partitioning around medoids is a variant of K-means which offers the advantage being more robust to noise and outliers. As the name suggests it uses an actual point in the cluster to act as medoid. A medoid is the centrally located object in the cluster with minimum sum of distances to other points. The presence of outlier in data greatly affects the mean of the cluster, whereas the medoid is unaffected by the outlier. The difference between K - means and K - medoids is similar to the difference between mean and median: where mean indicates the average value of all data items collected, while median indicates the value around that which all data items are evenly distributed around it. After finding the set of medoids, each object of the data set is assigned to the nearest medoid.

Cluster validation [33] is required to ascertain whether the number of clusters obtained by applying the aforementioned algorithms is indeed correct. The two criteria that govern the cluster validity indices available are: compactness/withinness and separation/between. Compactness measures the closeness of data points in a cluster, also called intra class similarity. Separation measures the distinction between clusters, also called inter class similarity. Clustering result is acceptable when a cluster possesses high intra class similarity and low inter class similarity. Some of the most commonly used cluster validity indices are: Elbow method using WCSS, Dunn Index, DB Index, Silhouette index.

WCSS is defined as Within-Cluster-Sum-of-Squares which is given by the formula (1):

$$WCSS = \sum_{i=1}^{n} (X_i - Y_i)^2 \tag{1}$$

where Y_i is the centroid corresponding to observation X_i, and *n* is the number of features used to describe the observation.

Elbow-Method using WCSS is one of the most common used methods which is based on the principle that while clustering performance as measured by *WCSS* increases (i.e.

WCSS decreases) with increase in k, rate of increase is usually decreasing. Plotting *WCSS* against increasing k can show an 'elbow' which demarks significant drop in rate of increase. Selecting number of clusters corresponding to elbow point achieves reasonable performance.

The Elbow-Method using WCSS was run by iteratively varying the number of clusters. The experiment validated the existence of 3 clusters in both population 1 and population 2.

Dunn index uses the minimum pairwise distance between objects in different clusters as the inter-cluster separation and the maximum diameter among all clusters as the intra-cluster compactness. The Silhouette index validates the clustering performance based on the pairwise difference of between- and within-cluster distances. In addition, the optimal cluster number is determined by maximizing the value of this index. The DB index can be obtained by averaging all the cluster similarities. The smaller the index is, the better the clustering result is.

Label validation

We wanted to further validate the class labels and so used the concept of Quality of experience to compute the MOS. The MOS given in formula (2) is calculated as the arithmetic mean over single ratings performed by individuals for a given stimulus in a subjective quality evaluation test. Thus:

$$MOS = \frac{\sum_{n=0}^{N} R_n}{N} \tag{2}$$

where R are the individual ratings for a given stimulus by N subjects. This score gives the agreeability of the obtained value with the actual value given by the individual.

Classifiers

Classification is a supervised machine learning technique that identifies data as belonging to a particular class or group based on a set of features which is discussed in Aggarwal et al. [34]. Different classification algorithms such as support vector machines, decision trees, naïve bayes classifier, K-nearest neighbour classifier and logistic regression are used to classify the target group.

Support Vector machine is a linear classifier that separates samples of different classes by using a decision boundary called hyperplane. SVMs can be used to classify both linear and non-linear data [35]. Support Vector machines have been extensively used in the classification of handwritten digits [36], classification of cancer tissue samples using microarray expression data [37], and text categorization [38]. So, SVMs is a technique that is proven to work well with concise and categorical data.

Decision tree is a hierarchical classifier that uses a set of rules to segment the predictor space. A decision tree has two types of nodes, leaf nodes and internal nodes. Each internal node represents a test on an attribute depending on which child nodes branch out. Leaf nodes have a class label determined by majority of the training samples reaching that leaf. Some applications of decision tree classification include classification of land cover from remotely sensed data Friedl et al. [39], diagnosis of ovarian cancer from mass spectral data Vlahou et al [40], classification of Alzheimer's disease based on MRI scans Zhang et al. [41] etc.,

K nearest neighbor is an instance based learning classifier. In this model, the new datapoint, is compared with k nearest sample datapoints, and the class with maximum number of nearest neighbors to the new datapoint is deemed as the class of the datapoint. This is called a lazy classification technique as nothing is done during the training phase other than just storing the training samples [42] Jiang et al.. This algorithm takes up a lot of memory and usually works well with less number of dimensions. Some applications of K-nearest neighbor classification include intrusion detection Liao et al. [43], classification of handwritten digits Lee et al [36] etc.

Naive Bayes Classifier is a statistical classifier that predicts class membership probabilities such as the probability that a given tuple belongs to a particular class. It is based on Bayes theorem. The naive assumption of class-conditional independencies made. This presumes that the attributes' values are conditionally independent of one another, given the class label of the tuple. Liu et al [44] performed sentiment analysis on a large dataset using a specially implemented naïve bayes classifier and achieved an accuracy of 82%. They observed that the accuracy improved as the size of the dataset increased.

Logistic Regression is a probabilistic linear classifier. The idea of logistic regression is to make linear regression produce probabilities. It's always best to predict class probabilities instead of predicting classes. Logistic regression estimate class probabilities directly using the logit transform. It is used for converting class probabilities into classes. It takes in the probability value and outputs either 0 or 1 as class label for binary classification problems S. Dreiseitl et al. [45]. The work by Ribeiro et al [46] clearly explains the concept of trust in choosing between classifiers. It also suggests the use of the most appropriate features to enhance the trust score of classifier.

An ensemble is a composite model made of a combination of classifiers. Bagging is an ensemble method in which each classifier votes and the class label prediction returned by the ensemble is based on these votes. Given a new data tuple to classify, the base classifiers each vote by returning a class prediction. The ensemble returns a class prediction based on the votes of the base classifiers. Boosting is an ensemble method which is incrementally built by training each subsequent model to emphasize more on the training samples that were misclassified by the previous models Kuncheva et al. [47].

Random Forest is a Tree Ensemble that operates by constructing a number of decision trees at training time and outputting the class that is the mode of the classes of the individual trees Brieman Leo [48]. Some applications of random forests include land cover classification. Gislason et al [49], image classification and micro array based cancer classification Statnikov et al [40].

Classifier performance measures

The choice of classifier for the data in hand is based on the performance measures. The confusion matrix is the foundation for computation of the performance measures. The evaluation of the optimal solution during the classification training can be defined based on confusion matrix. The row of the table represents the predicted class, while the column represents the actual class. From the confusion matrix given in Table 1, tp and tn denote the number of positive and negative instances that are correctly classified. Meanwhile, fp and fn denote the number of misclassified negative and positive instances, respectively.

The performance measure of the various classifiers are evaluated using the accuracy, precision, recall and f-1 scores. The values of tp, tn, fp, fn, P and N are used in the calculation of Accuracy, Precision and Recall. The formulae for calculating the aforementioned performance measures are given in Table 2.

Accuracy or recognition rate is the number of correctly classified test samples to the total number of test samples. Average accuracy for each class gives the accuracy of the entire classifier.

Precision can be thought of as a measure of exactness (i.e., what percentage of tuples labelled aspositive are actually such). The precision is intuitively the ability of the classifier not to label as positive a sample that is negative. The precision is the ratio tp /(tp + fp) where tp is the number of true positives and fp the number of false positives.

Recall is a measure of completeness (what percentage of positive tuples are labeled as such). Recall is intuitively the ability of the classifier to find all the positive samples. The recall is the ratio tp / (tp + fn) where tp is the number of true positives and fn the number of false negatives.

The F-beta score can be interpreted as a weighted harmonic mean of the precision and recall, where an F-beta score reaches its best value at 1 and worst score at 0.

Table 1 Confusion matrix

Actual value			
Predicted value		Positive	Negative
	Positive	tp	fp
	Negative	fn	tn

Table 2 Performance measures of a classifier

Performance measure	Formula
Accuracy	$\frac{TP+TN}{P+N}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F-Score	$\frac{2^{*}precision^{*}recal}{precision+recall}$

Experiments

The questionnaire comprises of 20 questions which was given to 300 individuals in population 1 and 356 individuals in population 2. The target population 1 were aged between 18 to 21 years old, whereas the target population 2 were aged between 22 to 26 years old. The 20 questions are considered as features and the responses collected are considered as data points. Each question had 5 different answers to choose from namely: almost never, sometimes, often, very often and almost always. The weights given for answers vary from 1 to 5 for the choices mentioned above.

Data preprocessing - clustering and label validation

The purpose of clustering in this research work is to identify the possible groups in the two target population taken under consideration. The 300 samples and 356 samples were collected from population1 and population 2 respectively. These samples are subject to various clustering algorithms like K-Means clustering, hierarchical clustering and partitioning around medoids or K- medoids.

K-means clustering uses an iterative heuristic approach to find the number of clusters K. It takes as input the data points i.e., the number of samples and the number of clusters K. Repeated experiments were run by varying the values of K. The results of the runs executed were recorded for further processing to ascertain the number of clusters K in the two populations under consideration.

We have used the agglomerative approach of hierarchical clustering. It is a bottom up approach that initially considers every data point as a cluster and then finds similarity between data points by using the Euclidean distance method. The 300 and 356 data samples of population 1 and population 2 were respectively subject to hierarchical clustering. The input for hierarchical clustering is the data samples and the distance to merge clusters in the agglomerative approach.

Clustering using K – Medoids or Partitioning Around Medoids (PAM) takes as input the data samples from population 1 and population 2 along with the heuristic for number of clusters K. The algorithm takes as centroid and actual data point and iteratively computes the distance between adjoining

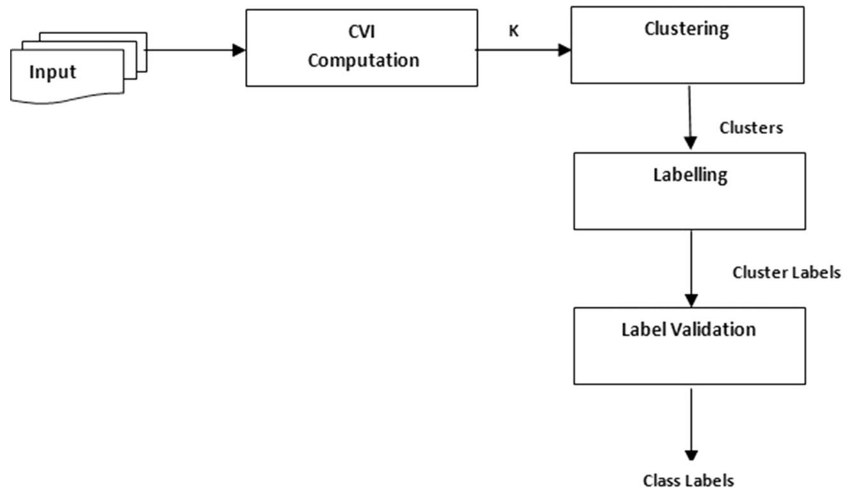


Fig. 1 Steps in data preprocessing

data points depending on the number of clusters K given as input for that iteration.

In order to validate the labels obtained by clustering we have used the concept of MOS. As a part of this process we communicated the class label to each of the individuals who participated in the experiment to find the individual agreeability with their state of mental health.

The appropriate number of clusters and its corresponding labels were identified. The next step is to build a classifier models with the features and the labels obtained through clustering. The target population 1 and population 2 resulted in the same number of clusters, intuitively we have decided to merge them into a single population with 656 samples.

The Fig. 1 shows the steps in data preprocessing, which explains the steps involved to identify the number of clusters and also to determine the class labels to build prediction models.

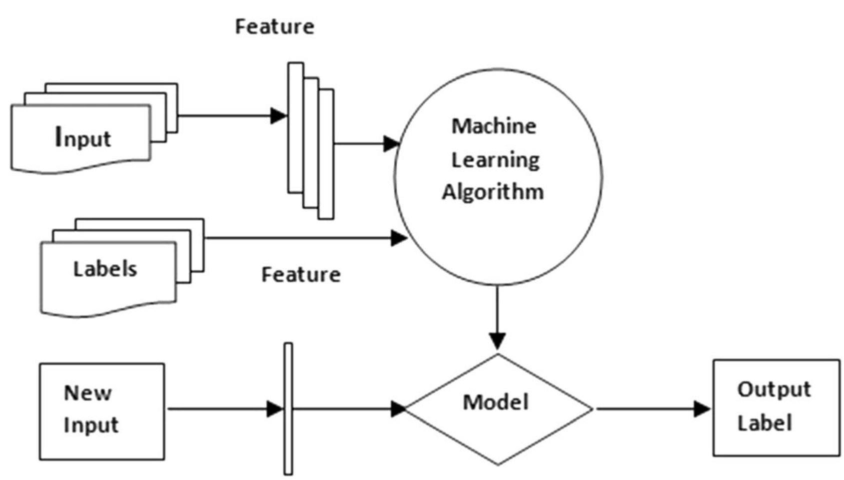
Classification

Our dataset comprises of 656 samples, 20 features, 3 class labels. The mean age of the population under consideration is 22 years. The training set and test set was fixed as 80:20 respectively. The training set comprises of 525 samples and training set comprises of 140 samples. With this corpus we proceed further to build the classifier models using different classification algorithms. The classification algorithms used are: Logistic regression, Naïve bayes, Support vector machines, Decision tree and K-nearest neighbors. In addition to this we have also built models using ensemble bagging method and tree ensemble using random forest.

The training set is fed into multiple classifiers and the models are built. The performance of each classifier is evaluated using the test set. The performance was measured using precision, recall, f-1 and accuracy measures. To improve the accuracy of the classification, ensemble methods were used. Bagging was implemented using Logistic regression, Support vector machines, Decision tree and K-nearest neighbors. A tree ensemble was also implemented using the random forest classifier.

The generic steps used in the construction of classifiers used different classifier models are given in Fig. 2. The performance measures of classifiers measure the decision making capability of the classifier. The measures used to determine the

Fig. 2 Construction of classifier



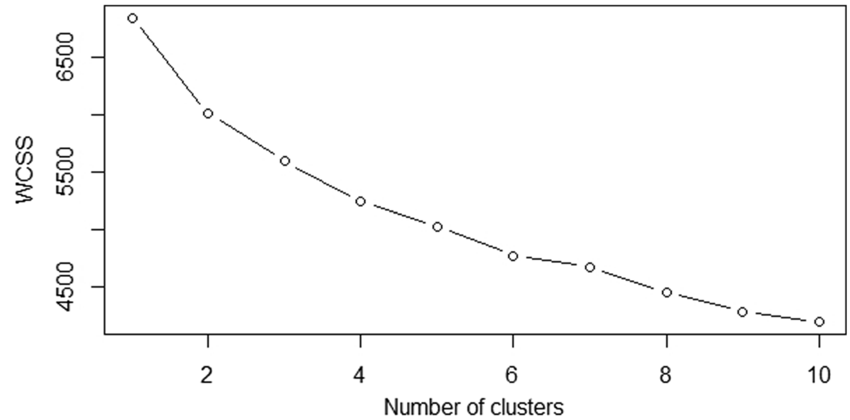


Fig. 3 Elbow diagram for target population 1

performance are: accuracy, precision, recall, F-score. The overall effectiveness of classifier was computed by the accuracy score. The precision, recall, F-score of the class labels across the classifiers are compared to decide the optimal classifier model. The trust of the classifier was also experimented using the LIME method proposed in [46]. The inference from LIME is that it works best when a subset of features is used for trust computation using classifiers.

Results and discussions

The research was initially focused on two target populations namely population 1 in the age group 18 to 21 years and the population 2 in the age group 22 to 26 years. Clustering was considered as the best choice for identifying groups in population 1 and population 2. The clustering algorithms used are: K-Means Clustering, hierarchical clustering and K- Medoids.

In K-Means clustering, repeated experiments were run by varying the values of K. Here first the algorithm is run with 300 population 1 data samples and the results were recorded for different values of K varying between 1 to 10. The output of the K-Means clustering for K = 3 was better than the other iterations.

The 300 and 356 data samples of population 1 and population 2 were respectively subject to hierarchical clustering. The distance is varied for 10 consecutive runs to determine the number of clusters. There was minimum overlap between clusters when the cluster size was 3. So the output of hierarchical clustering is 3 clusters.

The K-Medoids clustering is run with 300 and 356 data samples of population 1 and population 2 respectively by given an actual data point as centroid. The centroid value is varied and iteratively the results were recorded for 10 runs. The output of K-medoids clustering also resulted in 3 clusters.

Cluster validation is used to ascertain the actual number of clusters K is correct. We have used elbow method using WCSS to validate the value of K.

We have used WCSS method and elbow method validate the number of clusters obtained. The validation of cluster produces the output as 3 for both the samples of population 1 and population 2.

The illustration of the elbow method applied for population 1 is shown in Fig. 3 and for population 2 is shown in Fig. 4.

The number of clusters obtained for target population 1 is 114,94,92 and for target population 2 is 146,104,106. The next step is to determine the class labels for the 3 clusters obtained for population 1 and population 2.

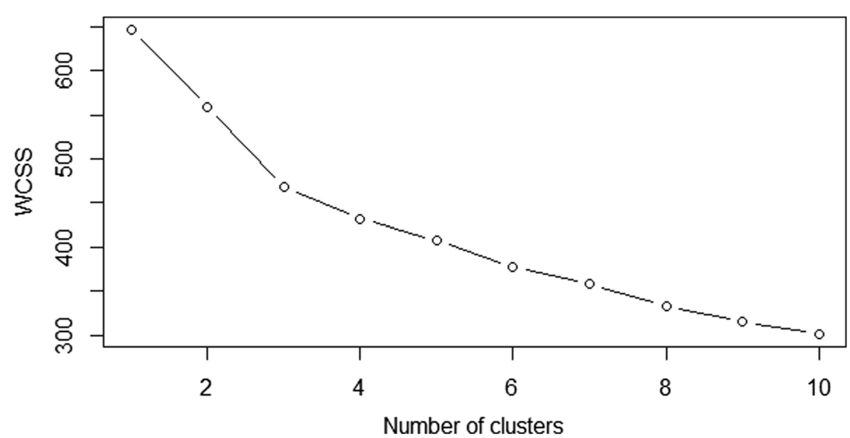


Fig. 4 Elbow diagram for target population 2

Table 3 Accuracy of classifiers

Classifier	Accuracy
Logistic Regression	0.84
Naïve Bayes	0.73
Support Vector Machine	0.89
Decision Tree	0.81
K Nearest Neighbour	0.89
Ensemble (Bagging)	0.90
Tree Ensemble (Random Forest)	0.90

The choice of answers were given a scale of 1 to 5, where 1 being the least (almost never) and 5 being the highest (almost always). The scores for each respondent in both the target populations were computed. The score of 0–30 was labeled as mentally distressed, score of 31–60 was labeled as neutral and score 61–100 was labeled as happy. The similarity of the scores obtained from both the target population give the idea to build a classifier considering the samples of both the population as one large sample with 656 responses.

The labels obtained were further validated using the concept of MOS. The individuals who were a part of target population were sent the responses i.e., the labels to find whether they agree with the finding. Here also they were given a scale of 1 to 5 as response to the finding. A scale of 1 means disagree and 5 being totally agree. The scores obtained for mentally distressed, neutral and happy were 4.67, 4.17 and 3.86 respectively.

The performance measures of classifiers measure the decision making capability of the classifier. The measures used to determine the performance are: accuracy, precision, recall, F-score.

The overall accuracy measure is rarely considered adequate for a classifier. But still it is necessary to have a notion of the correctly classified samples in the test set. It gives an overall effectiveness of a classifier. The accuracy scores obtained for the classifiers built are given In Table 3. It is clear that SVM, KNN, ensemble (bagging) and tree ensemble (random forest) give an equivalent accuracy score of 0.9.

We have to calculate the other performance measures before deciding on a suitable classifier for our dataset.

Table 5 Recall scores of each classifier

Classifier	Mentally distressed	Neutral	Happy
Logistic Regression	0.94	0.78	0.80
Naïve Bayes	0.9	0.47	0.83
Support Vector Machine	0.9	0.89	0.88
Decision Tree	0.88	0.71	0.83
K Nearest Neighbour	0.92	0.93	0.80
Ensemble (Bagging)	0.92	0.91	0.86
Tree Ensemble (Random forest)	0.9	0.91	0.88

The precision is a measure of the class agreement of the data labels with the positive labels given by the classifier. Since it is directly to class labels we have to calculate the precision scores for each of the 3 class labels. The values for precision score for each of the classifiers along with the 3 labels used in this research are given in Table 4. We are more interested in targeting people who are distressed because they are the ones who will need attention more than others. We can see that the classifiers SVM, KNN, ensemble (bagging) and tree ensemble (random forest) give a score of 0.95 for the mentally distressed class label. The score close to 1 indicates that the data sample labeled as mentally distressed does really belong to that class label.

Recall also known as sensitivity is the measure represent the effectiveness of the classifier to identify class labels. Here also we are focused on getting a score close to 1 for the mentally distressed class label. The recall scores for 3 class labels and the classifiers are shown in Table 5. The recall score 0.95 was obtained for the mentally distressed class label for the classifiers: SVM, KNN, ensemble (bagging) and tree ensemble (random forest).

F- Score gives the relationship between positive labels and those given by the classifier. It is computed by taking the harmonic mean of precision and recall for all the 3 labels across all the classifiers. The score close to 1 for the mentally distressed class label is desirable for deciding the best model of classifier. The F scores for the class labels are shown in Table 6. Again the classifiers SVM, KNN, ensemble (bagging) and tree ensemble

Table 4 Precision scores of each classifier

Classifier	Mentally distressed	Neutral	Happy
Logistic REGRESSION	0.87	0.83	0.82
Naïve Bayes	0.70	0.70	0.81
Support Vector Machine	0.95	0.87	0.84
Decision Tree	0.89	0.78	0.73
K Nearest Neighbour	0.93	0.82	0.93
Ensemble (Bagging)	0.95	0.85	0.88
Tree Ensemble (Random forest)	0.95	0.85	0.88

Table 6 F score for each classifier

Classifier	Mentally distressed	Neutral	Happy
Logistic Regression	0.90	0.80	0.81
Naïve Bayes	0.78	0.57	0.82
Support Vector Machine	0.92	0.88	0.86
Decision Tree	0.88	0.75	0.77
K Nearest Neighbour	0.92	0.87	0.86
Ensemble (Bagging)	0.93	0.88	0.87
Logistic Regression	0.90	0.80	0.81

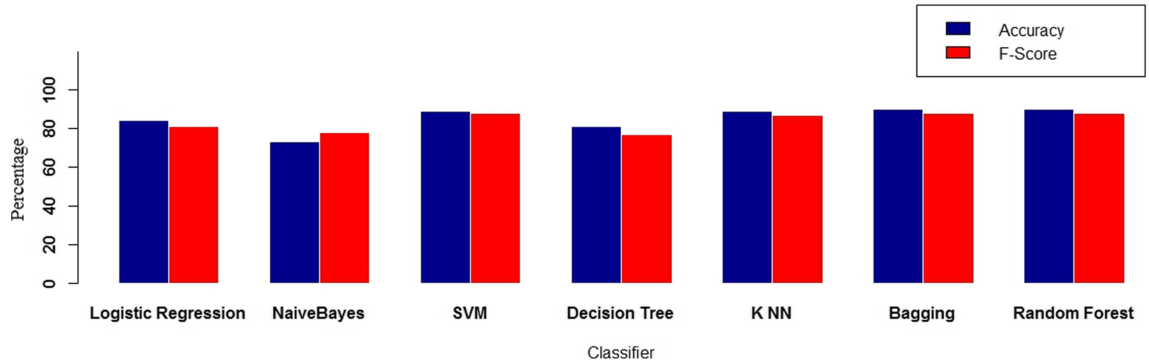


Fig. 5 Overall Comparison of Accuracy and F-Score across Classifiers

(random forest) have clearly become the contenders to decide the optimal classifier for our dataset.

The objective of predicting the mentally distressed individuals in a given target group with higher precision is achieved by SVM, Ensemble and Random Forest. Among the classifiers SVM and KNN perform with reasonable accuracy. As can be seen from the tables, support vector machines and K-Nearest neighbours give the best performance as individual classifiers. Support vector machines work well for the problem as this is concise set of data and because the labels are weakly known. K-Nearest neighbour works well as there are lesser number of dimensions or features. Naïve bayes does not give a good performance as the assumption of class conditional independence will work only for a huge dataset. Although logistic regression avoids overfitting and is robust to noise, it requires strong correlation between the X and Y variables and hence, it provides a fair performance, if not exceptional. Decision tree does not work too well as it not robust to noise and does not generalize well future observed data. An ensemble tends to be more accurate than its base classifiers. Ensembles yield better results when there is significant diversity among the models. A tree ensemble was implemented using the random forest classifier and an adhoc bagging ensemble was implemented using an ensemble of logistic regression, support vector machines, Decision trees and K-nearest neighbour as it empirically provided the best performance. The overall performance comparison is shown in Figs. 5 and 6.

Conclusion and future work

This paper provides an intuitive understanding of the mental health analysis amongst different target groups. We have developed a framework for determining the state of mental health of an individual. This framework was used to build prediction models. Prior to building models, clustering algorithms were used identify the number of clusters. The class labels obtained were validated using MOS, which were given as inputs to train the classifier. The experiments have demonstrated that SVM, KNN, Random Forest have performed almost equivalently. Also the use of ensemble classifiers was found to significantly improve the performance of the mental health prediction with 90% accuracy. This work can be extended to include different sections of the society and also categorizing different mental illness like anxiety, depression, etc., From the results obtained we feel that the workflow suggested here can be used a mechanism to perform behavioral modeling of a target population. The use of physiological parameters like electrocardiogram (ECG), respiratory rate could also be included as features in order to predict the mental state of an individual more appropriately. The inclusion of physiological parameters also adds to the increase in the number of features in the dataset. The interpretation of physiological values and arriving at features could itself be a challenging task. The quality of the features directly has an impact on the

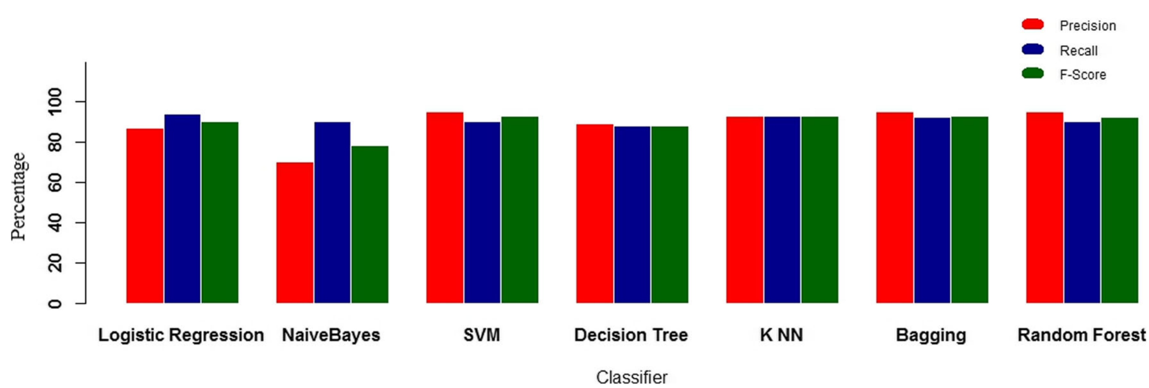


Fig. 6 F-Score, Precision and Recall for Mentally Distressed class of people

reliability of the prediction models built. An estimate of the contribution of features can be used to determine the optimal number of features to be used for building the model. Feature subset selection strategies could then be used to select the appropriate features. This would reduce the time taken to build prediction models. The decision of choosing between different models can be done and validated using the LIME package. The framework suggested here could also cater to a wide range of mental illness by including the concept of fuzziness in building models. When there are more categories of mental illness in the target population, naturally it will result more number of class labels. In such cases there tends to be more overlap between some class labels. This problem can be addressed by writing fuzzy rules in cases where the overlap is expected. The concept of Deep Learning can be used for very large dataset. The classification accuracy can also be improved using deep learning methods such as recursive neural networks. It also enables us to cater to much wider community which will result in more data samples.

Compliance with ethical standards

Ethical approval "All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."

Informed consent "Informed consent was obtained from all individual participants included in the study."

References

1. Miner, L., et al., *Practical predictive analytics and decisioning systems for medicine: Informatics accuracy and cost-effectiveness for healthcare administration and delivery including medical research*. Cambridge: Academic Press, 2014.
2. Luxton, D. D., (ed.) *Artificial Intelligence in Behavioral and Mental Health Care*. Amsterdam: Elsevier Inc., 2015.
3. Hahn, T., Nierenberg, A. A., and Whitfield-Gabrieli, S., Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol. Psychiatry* 22(1):37–43, 2017.
4. Bijl, R. V., Ravelli, A., and Van Zessen, G., Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc. Psychiatry Psychiatr. Epidemiol.* 33(12):587–595, 1998.
5. World Health Organization, Mental health: a call for action by world health ministers. Geneva: World Health Organization, Department of Mental Health and Substance Dependence, 2001.
6. Funk, M., Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level. http://apps.who.int/gb/ebwha/pdf_files/EB130/B130_9-en.pdf. Accessed 20 Feb 2016, 2016.
7. Drapeau, A., Marchand, A., and Beaulieu-Prévost, D. Mental illnesses-understanding, prediction and control. *Epidemiol. Psychol. Distress*, (2012). https://doi.org/10.5772/1235.
8. Goodman, R., Renfrew, D., and Mullick, M., Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *Eur. Child Adolesc. Psychiatry* 9(2):129–134, 2000.
9. Lanata, A. et al., Complexity index from a personalized wearable monitoring system for assessing remission in mental health. *IEEE J. Biomed. Health Inform.* 19(1):132–139, 2015.
10. Schaefer, J. D., et al. "Enduring mental health: Prevalence and prediction.". *J. Abnorm. Psychol.* 126(2):212, 2017.
11. Qiu, T., Zhang, Y., Qiao, D., Zhang, X., Wymore, M. L., & Sangaiah, A. K., A robust time synchronization scheme for industrial internet of things. *IEEE Trans. Ind. Inf.*, 2017. https://doi.org/10.1109/TII.2017.2738842.
12. Qiu, T., Qiao, R., Han, M., Sangaiah, A. K., and Lee, I., A Lifetime-Enhanced Data Collecting Scheme for the Internet of Things. *IEEE Commun. Mag.* 55(11):132–137, 2017.
13. Kumar, P., Kumari, S., Sharma, V., Sangaiah, A. K., Wei, J., and Li, X., A Certificateless aggregate signature scheme for healthcare wireless sensor network. *Sustain. Comput. Inf. Syst.*, 2017. https://doi.org/10.1016/j.suscom.2017.09.002.
14. Sangaiah, A. K., Samuel, O. W., Li, X., Abdel-Basset, M., and Wang, H., Towards an efficient risk assessment in software projects–Fuzzy reinforcement paradigm. *Comput. Electr. Eng.*, 2017. https://doi.org/10.1016/j.compeleceng.2017.07.022.
15. Wu, F., et al., A lightweight and robust two-factor authentication scheme for personalized healthcare systems using wireless medical sensor networks. *Futur. Gener. Comput. Syst.* 82:727–737, 2017.
16. Aborokbah, M. M., Al-Mutairi, S., Sangaiah, A. K., and Samuel, O. W., Adaptive context aware decision computing paradigm for intensive health care delivery in smart cities—A case analysis. *Sustain. Cities Soc.* 2017.
17. Chinaveh, M., The effectiveness of problem-solving on coping skills and psychological adjustment. *Procedia. Soc. Behav. Sci.* 84:4–9, 2013.
18. Hajiyakhchali, A., The Effects of Creative Problem Solving Process Training on Academic Well-being of Shahid Chamran University Students. *Procedia. Soc. Behav. Sci.* 84:549–552, 2013.
19. Aghaei, A., Khayyamnekouei, Z., and Yousefy, A., General health prediction based on life orientation, quality of life, life satisfaction and age. *Procedia. Soc. Behav. Sci.* 84:569–573, 2013.
20. Strauss, J., Peguero, A. M., and Hirst, G., Machine learning methods for clinical forms analysis in mental health. *MedInfo.* 192:1024, 2013.
21. Jung, Y., and Yoon, Y. I., Multi-level assessment model for wellness service based on human mental stress level. *Multimedia Tools and Applications* 76(9):11305–11317, 2017.
22. Wang, H., and Wang, J., An effective image representation method using kernel classification. Tools with Artificial Intelligence (ICTAI), 2014 I.E. 26th International Conference on. IEEE, 2014.
23. Cheng, X. et al., iATC-mISF: a multi-label classifier for predicting the classes of anatomical therapeutic chemicals. *Bioinformatics* 33(3):341–346, 2016.
24. Rakesh, G., Suicide Prediction With Machine Learning. *Am. J. Psychiatry Residents' J.* 12(1):15–17, 2017.
25. Ribeiro, J. D. et al., Letter to the Editor: Suicide as a complex classification problem: machine learning and related techniques can advance suicide prediction-a reply to Roaldset (2016). *Psychol. Med.* 46(9):2009, 2016.
26. Kessler, R. C., et al., Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol. Psychiatry.* 21(10):1366, 2016.
27. Fleury, A., Vacher, M., and Noury, N., SVM-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results. *IEEE Trans. Inf. Technol. Biomed.* 14(2):274–283, 2010.

28. Smets, E., et al. Comparison of machine learning techniques for psychophysiological stress detection. *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer International Publishing, 2015.
29. Xu, J., et al. On the properties of mean opinion scores for quality of experience management. *Multimedia (ISM), 2011 I.E. International Symposium on*. IEEE, 2011.
30. Jung, Y. G., Kang, M. S., and Heo, J., Clustering performance comparison using K-means and expectation maximization algorithms. *Biotechnol. Biotechnol. Equip.* 28(sup1):S44–S48, 2014.
31. Kern, M. L., et al. "The EPOCH Measure of Adolescent Well-Being.". *Psychol. Assess.* 28(5):586, 2016.
32. Milligan, G. W., and Cooper, M. C., Methodology review: Clustering methods. *Appl. Psychol. Meas.* 11(4):329–354, 1987.
33. Dziopa, T., Clustering Validity Indices Evaluation with Regard to Semantic Homogeneity. FedCSIS Position Papers 2016.
34. Aggarwal, C. C., and Zhai, C. X., A survey of text classification algorithms. *Mining text data*. Springer US, 163–222, 2012.
35. Burges, C. J. C., A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Disc.* 2(2):121–167, 1998.
36. Lee, Y., Handwritten digit recognition using k nearest-neighbor, radial-basis function, and backpropagation neural networks. *Neural Comput.* 3(3):440–449, 1991.
37. Statnikov, A., Wang, L., and Aliferis, C. F., A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics* 9(1): 319, 2008.
38. Joachims, T., Text categorization with support vector machines: learning with many relevant features. In: *European conference on machine learning*. Pp. 137–142. Berlin, Heidelberg: Springer, 1998.
39. Friedl, M. A., and Brodley, C. E., Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* 61(3): 399–409, 1997.
40. Vlahou, A. et al., Diagnosis of ovarian cancer using decision tree classification of mass spectral data. *Biomed. Res. Int.* 2003(5):308–314, 2003.
41. Zhang, Y., Wang, S., and Dong, Z., Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Prog. Electromagn. Res.* 144: 171–184, 2014.
42. Jiang, L., et al. Survey of improving k-nearest-neighbor for classification." *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*. Vol. 1. IEEE, 2007.
43. Liao, Y., and Rao Vemuri, V., Use of k-nearest neighbor classifier for intrusion detection. *Comput Secur* 21(5):439–448, 2002.
44. Liu, B., et al. Scalable sentiment classification for big data analysis using naive bayes classifier. *Big Data, 2013 I.E. International Conference on*. IEEE, 2013.
45. Dreiseitl, S., and Ohno-Machado, L., Logistic regression and artificial neural network classification models: a methodology review. *J. Biomed. Inform.* 35(5):352–359, 2002.
46. Ribeiro, M. T., Singh, S., and Guestrin, C., Why should i trust you?: Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.
47. Kuncheva, L. I. *Combining pattern classifiers: methods and algorithms*. New York: John Wiley & Sons, 2004.
48. Breiman, L., Random forests. *Mach. Learn.* 45(1):5–32, 2001.
49. Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R., Random forests for land cover classification. *Pattern Recogn. Lett.* 27(4): 294–300, 2006.