International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-5, Issue-2, June 2016

Classifying Datasets Using Some Different Classification Methods

Sanaa Abou Elhamayed

***Abstract*— The classification methods are used to classify a new data instance based on the known classifications of the observations in the training set. The main objective of this work is to compare the performance of three classification methods. The methods are called the Classification and Regression Tree (CART), K-Nearest Neighbour, (KNN), and Principal Component Analysis (PCA). Such methods are applied on different datasets. Any dataset is partitioned into two sets, one of them is training set and the other one is testing set. The performance of each method is measured using some measurable criteria. This includes: non-error rate, error rate, accuracy, precision, sensitivity, and specificity. The adopted methods are evaluated and compared using some chosen datasets as testbeds. The cross validation is applied to improve and assesst the performance of the classification methods. The classification methods are implemented and operated by applying MATLAB version-4 for calculating the significant parameters which have a direct effect on the performance of the classification methods.**

***Index Terms*— Classification Methods, CART, KNN, PCA, Cross Validation, Qualification Parameters**

I. INTRODUCTION

A lot of research works have been presented concerning the classification problems. This involves using classification algorithms, software tools, datasets, and classification accuracies. Examples of such published efforts are briefly mentioned as follows:

[1] conducted an experiment using the WEKA environment by handling four classification algorithms namely ID3, J48, Simple Classification And Regression Tree (CART) and Alternating Decision Tree on the spam email dataset. Such classification algorithms are used to categorize the emails as spam or non-spam. The algorithms are analyzed and compared in terms of classification accuracy. From the results it was found that the highest accuracy performance is for the J48 classifier for the spam email datasets containing 4601 instances with58 attributes per each.

[2] used decision tree classification algorithms to classify the data into correctly and incorrectly instances. Their work shows the process of WEKA analysis and selection of attributes to be mined. Also, they provided an evaluation based on the

evolutionary classification algorithms to their datasets and measured the accuracy of the obtained results.

[5] presented an introduction of text classification and compared some existing classifiers according to time complexity, principal, and performance. They verified that information Gain and Chi square statistics are the most commonly used and well performed methods for feature selection. Also, they verified that no single representation scheme and classifier can be mentioned as a general model for any application. Different algorithms perform differently depending on data collection.

[6] presented a comparative study of different classification and clustering techniques using WEKA. They tested J48, ID3, Bayes network classification algorithms. According to their comparison they verified that J48 algorithm gives the best performance considering both accuracy and speed.

[7] presented some classification techniques which are decision tree, Bayesian networks, k-nearest neighbour classifier, neural network, and support vector machine. These techniques are used to uncover hidden patterns within large amounts of data and predict their future behaviour. They verified that the good data is the first requirement for good data exploration.

[8] evaluated the performance of data mining classification algorithms on various datasets. They found that most algorithms can classify datasets with both nominal and numeric class values. But bayes algorithms classify datasets with only nominal class values whereas linear regression, M5 rules classify datasets only with numeric class value. They found that J48 algorithm performed well with 100% correctly classified instances with least time.

[9] applied five different classification methods for classifying different types of data based on their size. The five classification methods are decision tree, lazy learner, rules based, naive bayes, and regression. They proposed the data using WEKA tool which provides working with attributes section and evaluate the performance of the classification algorithms according to the accuracy and the error rate. They found that the lazy learner is much better than the others in big datasets while the rules basedis good in small datasets. They also found that the decision tree does not change when a dataset is changed.

[10] analyzed the performance of three Meta classification algorithms namely attributed selected classifier, filtered classifier and logitboost. They analyzed the performance of the algorithms by evaluating the classification accuracy and error rate. They classified the computer files according to their extension and used the WEKA tool for analyzing the performance of the classification algorithms. The dataset is collected from the computer systems and contains 9000 instances and four attributes. Before starting the classification process the training and the test data are reduced by attribute

Sanaa Abou Elhamayed, Electronic Research Institute, Giza, Egypt

selection. From the experimental results it is observed that the logitbbost is better than the other algorithms.

[11] used two classification algorithms namely J48 and multilayer perceptron for several datasets for making a decision which is better based on the conditions of the datasets. The confusion matrix is used to evaluate the classification quality, where the sum of diagonal is the number of correctly classified instances else are incorrectly classified. They found that the multilayer perceptron is a better algorithm in the most of the cases.

[12] presented a study to find the best classification algorithm among bayesien and lazy classifiers. The dataset is collected from the computer files and has 80000 instances and four attributes. Bayesian algorithms predict the class depending on the probability of belonging to that class. Lazy algorithms predict the class depending on the distance from the test instance and its neighbours. They evaluated the quality of the classifying algorithms considering some measurable criteria such as: error rate, accuracy, F measure, Receiver operating characteristics, True positive rate, and kappa statistics. From the experimental results it is observed that the lazy classifiers' k-nearest neighbour is better than the other techniques.

The organization of this paper is as follows: Section 2 presents the chosen classification methods. Section 3 implements the different datasets and the effective parameters in the classification process while section 4 discusses the results. Section 5 concludes the whole work.

II. The Chosen Classification Methods

2.1 Classification Using CART Method

Decision tree is one of the most important knowledge representation methods which attempt to build a top-down method to reduce dimensionality. The reduction of dimensionality is used by eliminating duplicated or redundant attributes or neglecting less important ones. The decision tree method is used in different applications of science and medicine. Decision trees are trees that classify instances by sorting them based on features values. Each node in a decision tree represents a feature in an instance to be classified, each leave represents a class label, and each branch represents a conjunction of features that lead to those class labels (a value that a node can assume) [20], [23]. This method is based on rule induction. A distinction between continuous and categorical variables is required to describe the splitting rules. If the dataset has numerical variables then the number of possible splits at a given node is one less than the number of its distinctly observed values. If the dataset has M categorical variables then those variables will be splitted into M subsets. In this method the data set is recursively splitted into smaller subsets where each subset contains objects belonging to as few categories as possible. For the best splitting node, gain ratio and gini index are used. To decrease the height of the tree, the irregularity of each node must be reduced. So, the irregularity I is computed for all the features by applying $I = -\sum_c p(c) log_2 p(c)$ where p(c) is the proportion of the data that belongs to the class c. The final classification model consists of a tree that defines the classification rule.
The steps of the method can be summarized as follows:

Input: dataset (a set of feature vectors representing instances)

1. Create the root of the tree with the feature that maximizes the gain ratio G
$$G = I - \sum \frac{c_j}{c} I_c$$
2. Determine for the best split by computing the Gini index I_{gini}
$$I_{gini} = 1 - \sum_j p(c_j)^2$$
where p(c_j) is the relative frequency of cases belong to class c_j
3. Split the node into branches.
4. Check if branches have data.
5. Repeat.
6. Stop when all branches have no data.
7. Assign classes to terminal nodes.

Output: tree of classifying data and qualification parameters

2.2 Classification Using KNN Method

The k-nearest neighbour method is the most well known classification algorithm because of its simplicity. Also, it needs only two parameters to tune which are distance metric [10] observed that the k-nearest neighbour is better than the Bayseain algorithms.The k-nearest neighbour is called lazy classifier because it does not build a model until the time that a prediction is required. It only does work at the last second. Also, it is a competitive learning algorithm because it makes a comparison between data instances to make a predictive decision. The k-nearest neighbour algorithm predicts the unseen data instance by searching through the training dataset for the k-most similar instances. The prediction attribute of the most similar instances is summarized and returned as the prediction for the unseen instance. From training instance to sample instance distance is evaluated and the instance with lowest distance is called nearest neighbour. KNN method is used in many applications such as classification, problem solving, and function learning [25]. This method uses the Euclidean distance for the real valued data.

The steps of the algorithm can be summarized as follows:
Input: dataset (a set of feature vectors representing instances)

1. Specify a positive integer k.
2. Split the dataset into training dataset D and test dataset
D_z: the training dataset to make classifications and the test dataset to evaluate the accuracy of the algorithm.
3. Calculate the distance d(x',x) between the test instance z and every instance in the training dataset (x,y)Є D.
4. Select D_z D, the set of k closest training instances to test the instance z.
5. Find the most common classification of these instances (the majority class of the k nearest neighbours)

$$y' = argmax_v \sum_{(x_i,y_i)\epsilon D_z} I(v = y_i) \quad (2)$$

Where v is a class label, y_i is the class label for the i^{th} nearest neighbours, and i is an indicator function that

returns the value of 1 if its argument is true and 0 otherwise.

6. Give this classification to the test instance (the test instance is classified based on the majority class of its nearest neighbours).
7. Calculate the accuracy of the algorithm.
8. Collect the most similar all together.

Output: qualification parameters.

2.3 Classification Using PCA Method

Statistical procedure or learning uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The objective of PCA is to reduce the number of attributes (reduce the dimensionality) [24].

The steps of the algorithm can be summarized as follows:

Input: dataset (a set of feature vectors representing instances)

1. Compute the means of each attribute vector of all the data set by using the equation (3):

$$\bar{X} = \frac{\sum_{i=1}^{d} X_i}{d} \quad (3)$$

where $\bar{X}$ is the mean of the dataset X and d is the number of instances.

2. Subtract the mean from each of the data dimensions.
3. Compute the covariance matrix of the whole datasetX by using equation (4)

$$cov_{i,j} = \frac{\sum_{q=1}^{d} X_{q,i} . X_{q,j}}{d-1} \quad (4)$$

Where $cov_{i,j}$ is the covariance between attributes i and j.

4. Compute the eigenvectors for each attribute E = (e_1,e_2,-----,e_m) of the covariance matrix X is d×m .
5. Compute the corresponding eigenvalues Λ = diag (λ_1,λ_2,---,λ_d) where Xe=λe where λ is a scalar eigenvalue.
6. Sort the eigenvectors by decreasing the eigenvalues.
7. Get the eigenvectors with the largest eigenvalues to form the reduced matrix of dimension k.
8. Multiply the original matrix with the reduced one to form the matrix W with dimension d×k (transforms d×m matrix into d×k matrix).
9. Use the matrix W to transform the samples onto the new space.

Output: scattered samples figures and qualification parameters.

III. Implementation Work

To evaluate the performance of the adopted classification methods, four different datasets are used. The different datasets cover small dataset with large attributes, small data set with limited attributes, large dataset with limited attributes, and large dataset with large attributes. The datasets are Olitos, Glass, Diabetes, and Madelon datasets. Table (1) shows these datasets. Olitos dataset consists of 120 olive oil instances on measurements on 25 chemical compositions (fatty acids, sterols, triterpenic alcohols) of olive oils from Tuscany. There are 4 classes corresponding to different production areas. Class 1, Class 2, Class 3, and Class 4 contain 50, 25, 34, and 11 observations, respectively.

The Glass dataset consists of 214 glass samples of each 9 attributes which are: RI: refractive index, Na: Sodium, Mg: Magnesium, Al: Aluminum, Si: Silicon, K: Potassium, Ca: Calcium, Ba: Barium, and Fe: Iron. There are 7 classes corresponding to different types of glass which are :building_windows_float_processed, building_windows_non_float_processed,vehicle_windows_float_processed,vehicle_windows_non_float_processed, containers, tableware, headlamps.

The Diabetes dataset consists of 768 instances of each 8 attributes which are:

1. Number of times pregnant
2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-Hour serum insulin (mu U/ml)
6. Body mass index (weight in kg/(height in m)^2)
7. Diabetes pedigree function
8. Age (years)

Two class variables (0 or 1).

The Madelon is an artificial dataset which consists of 4400 instances of each 500 attributes

The Olitos dataset is divided randomly into training set (90 instances) and test set (30 instances). The glass dataset is divided randomly into training set with 144 instances and test set which 70 instances. The training and testing sets for the Diabetes are 568 instances and 200 instances and for the Madelon are 2600 instances and 1800 respectively.

Table 1: Datasets

Name	Instances	Attributes	Classes
Olitos	120	25	4
Glass	214	9	7
Diabetes	768	8	2
Madelon	4400	500	2

Cross validation is a popular strategy for method selection. The main idea of cross validation is to split data once or several times, for estimating the risk of each method: part of data (training sample) is used for training each method, and the remaining part (the test sample) is used for estimating the risk of method. Then, the cross validation selects the method with the estimated risk [Sylvain Arlot and Alian Celisse, 2010]. Software is the Classification toolbox for MATLAB - version 4.0 has been released by Milano Chemometrics and QSAR research Group. Visit their website at www.disat.unimib.it/chm. Hardware is Intel (R) Pentium 4, CPU 3.2 GHZ, and RAM 1.49 GB. The error rate is evaluated for the training set and test set.The hold out method is used to determine the stopping point. The hold out method used is cross validation. Random subsampling cross validation is applied to split the dataset randomly into training set and test set, then can calculate the error rate with the test. The qualification of the classification methods are based on the following parameters: non-error rate (NER) represents the average of the class sensitivity , error rate (ER), accuracy (Ac) is the ratio of correctly assigned samples , precision (Pr) is the ratio between the samples of g^{th} class correctly classified and the total number of samples assigned to that class, sensitivity

(Sn) describes the ability of the algorithm to recognize samples correctly, and specificity (Sp) characterizes the ability of the class to reject the samples of all other classes. These parameters are defined as in the following equations:

$$NER = \frac{\sum_{g=1}^{G} Sn_g}{G} \quad (5)$$

$$ER = 1 - NER \quad (6)$$

$$Ac = \frac{\sum_{g=1}^{G} n_{gg}}{n} \quad (7)$$

$$Pr_g = \frac{n_{gg}}{n'_g} \quad (8)$$

$$Sn_g = \frac{n_{gg}}{n_g} \quad (9)$$

$$Sp_g = \frac{\sum_{k=1}^{G}(n'_k - n_{gk})}{n - n_g} \quad \text{for } k \neq g \quad (10)$$

$$n'_k = \sum_{g=1}^{G} n_{gk} \quad (11)$$

Where :n'_g is the total number of samples assigned to the g^{th} class

n_{gg} is the number of samples belonging to class g and correctly assigned to it.

n_g is the total number of samples belonging to the g^{th} class.

$n'_k$is the total number of samples assigned to the k^{th} class.

IV. Discussion of Results

The quality of the classification methods are evaluated by precision, sensitivity, specificity, accuracy, non-error-rate, and error rate. Each element of the dataset is called an instance and the class it belongs to is called the label and the error rate of the dataset classifier is the probability of the classifier to incorrectly classify an instance. The chosen datasets are splitted into training set and testing set and the parameters that evaluate the quality of the chosen methods are shown in tables (2 to 9) respectively.

Table 2: Evaluation of the classifiers on Olitos Dataset

classifier	Non error rate	error rate	accuracy	Precision	sensitivity	specificity
CART	0.44	0.56	0.61	0.65	0.97	1
KNN	0.78	0.22	0.82	0.85	0.94	0.99
PCA	0.97	0.03	0.96	1	1	1

Table 3: Evaluation of the Classifiers on Olitos Test Dataset

classifier	Nonerror rate	error rate	accuracy	Precision	sensitivity	specificity
CART	0.25	0.75	0.7	0.7	1	1
KNN	0.3929	0.6071	0.7	0.9375	0.8571	1
PCA	1	0	1	1	1	1

Table 4: Evaluation of the Classifiers on Glass Dataset

classifier	Nonerror rate	error rate	accuracy	Precision	sensitivity	specificity
CART	0.8288	0.1712	0.8264	0.9016	0.9143	0.9143
KNN	0.8002	0.1998	0.7986	0.8462	0.8571	0.8571
PCA	0.7573	0.2427	0.7569	0.7746	0.7714	0.7714

Table 5: Evaluation of the Classifiers on Glass Test Dataset

classifier	Nonerror rate	error rate	accuracy	Precision	sensitivity	specificity
CART	0.5567	0.4433	0.9143	1	1	1
KNN	0.456	0.544	0.8	0.8929	0.9412	1
PCA	0.99441	0.0559	0.9	1	1	1

Table 6: Evaluation of the Classifiers on Diabetes Dataset

classifier	Nonerror rate	error rate	accuracy	Precision	sensitivity	specificity
CART	1	0	1	1	1	1
KNN	0.9936	0.0064	0.9947	0.9949	0.9973	0.9973
PCA	0.9936	0.0064	0.9947	0.9949	0.9973	0.9973

Table 7: Evaluation of the Classifiers on Diabetes Test Dataset

classifier	Nonerror rate	error rate	accuracy	Precision	sensitivity	specificity
CART	0.5	0.5	0.645	0.645	1	1
KNN	0.5016	0.4984	0.545	0.6462	0.6512	0.6512
PCA	0.5503	0.4497	0.665	0.6703	0.9457	0.9457

Table 8: Evaluation of the Classifiers on Madelon Dataset

Classifier	Nonerror rate	error rate	accuracy	Precision	sensitivity	specificity
CART	0.8615	0.1385	0.8615	0.8701	0.88501	0.8729
KNN	0.5606	0.4394	0.5605	0.5757	0.4635	0.6577
PCA	0.6105	0.3895	0.6105	0.6117	0.6074	0.6136

Table 9: Evaluation of the Classifiers on Madelon Test Dataset

Classifier	Nonerror rate	error rate	accuracy	Precision	sensitivity	specificity
CART	0.5	0.5	0.505	0.505	1	1
KNN	0.5021	0.4979	0.5028	0.5068	0.5699	0.5699
PCA	0.5362	0.4637	0.5367	0.5386	0.5754	0.5754

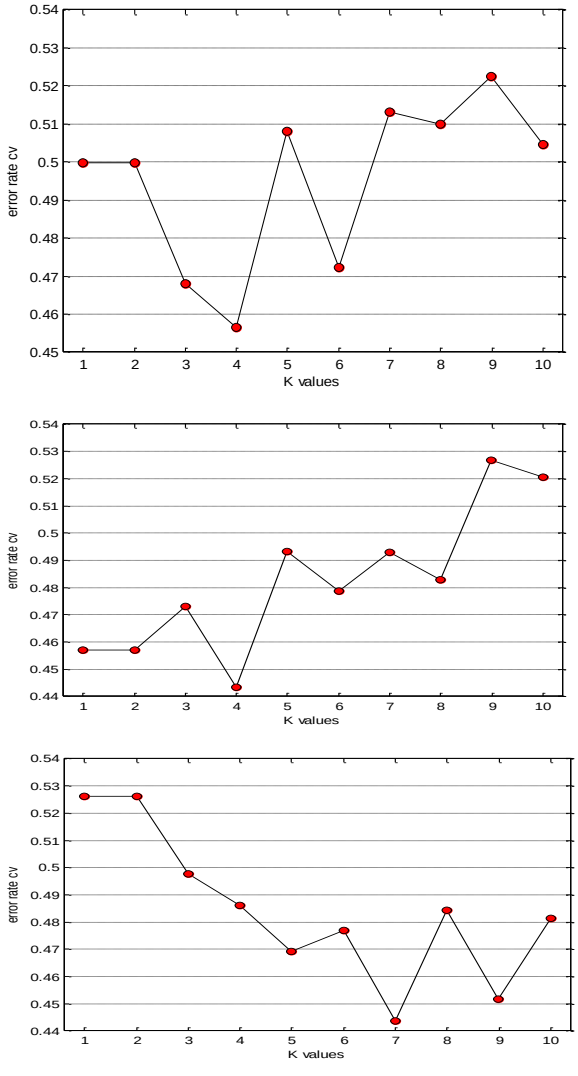
From the above tables, it is clear that when error rate increases accuracy, precision, sensitivity, and specificity decrease and vice versa for all classification methods using different datasets. From the test set, the classification algorithms are estimated to be applicable or not. If the accuracy, precision, sensitivity, and specificity of the algorithm are acceptable, the algorithm can be used to classify new data. Sensitivity and specificity are important statistical measures of the classification performance. Sensitivity measures the proportion of actual positives which are correctly identified as such. Specificity measures the proportion of negatives which are correctly identified. As shown from the tables for the same number of principle components, the sensitivity increases by increasing the specificity value. Although the specificity-sensitivity relationship is globally non-linear, it seems to be partially linear for some range values of specificity. Moreover, the specificity-sensitivity relationship changes by changing the number of principal components. The percentage accuracy changes by changing the number of principal components. Such changes may be in an increasing order, others in a decreasing order while others are alternating. If the dataset has large number of attributes, it is better to apply the PCA method. It is noticed that the KNN method is not sensitive for dataset has large attributes also takes more time than the other two methods. The PCA method is better than the other methods for big data but CART method is more sensitive. The runing time of the three

methods by applying the madelon dataset is shown in table (10).

Table 3: The Complexity Time of the Three Methods Using Madelon Dataset

Method	Time in Seconds
CART	50.2304
KNN	48.9867
PCA	5.9446

Because of the error rate of the training set is lower than the true error rate the dataset is partitioned in several different ways. The average score over the different partition is computed to avoid the possible bias introduced by relying on any particular division into test and train sets. The best method is estimated by how the method performs by the unknown data and the cross validation is used to measure the error rate using Diabetes dataset and applying PCA and KNN methods as shown in figures (1, 2, 3). The PCA aimed to finding the principle components with maximum dependence on the response variables. When the task is regression or classification, it is preferred to project the explanatory variables along directions that are related to the response variable. The two-dimensional projection results for the adopted dataset using the chosen method are shown in figure (3).



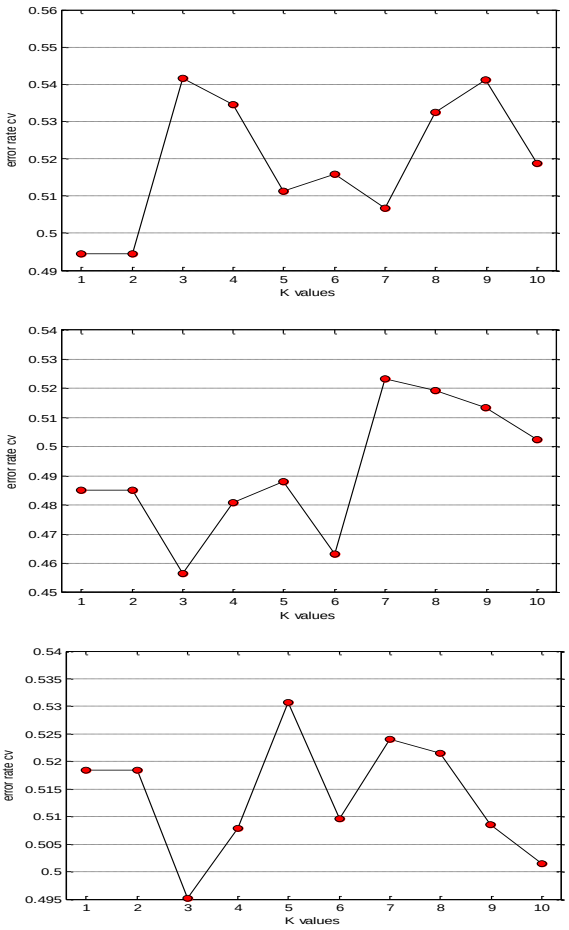
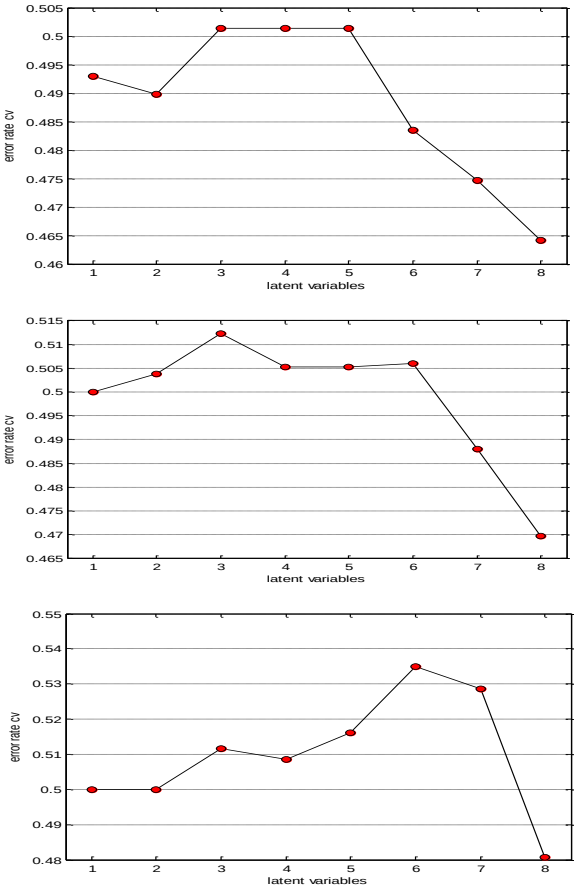


Fig. 1: Cross Validation Error Rate on Diabetes Dataset for the Different Segmentations 2, 3, 4, 5, 10, and all data as Computed by KNN Method



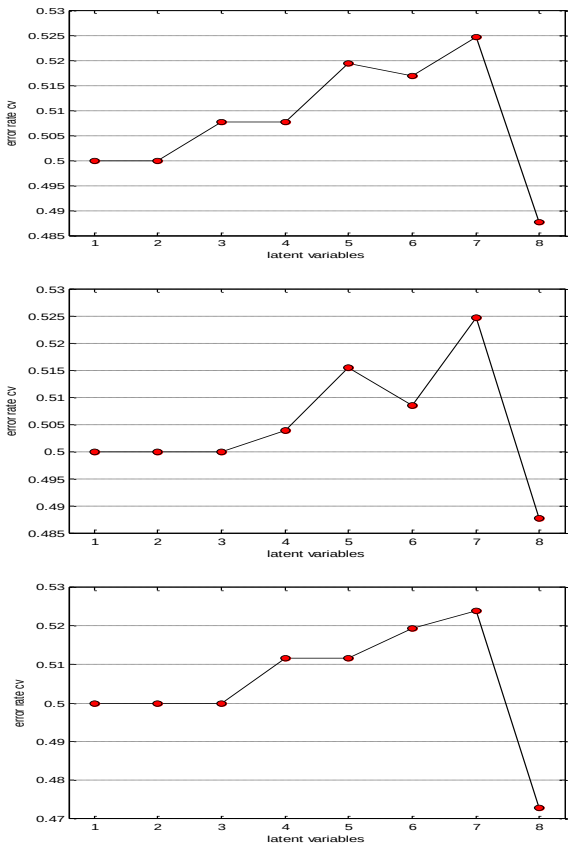
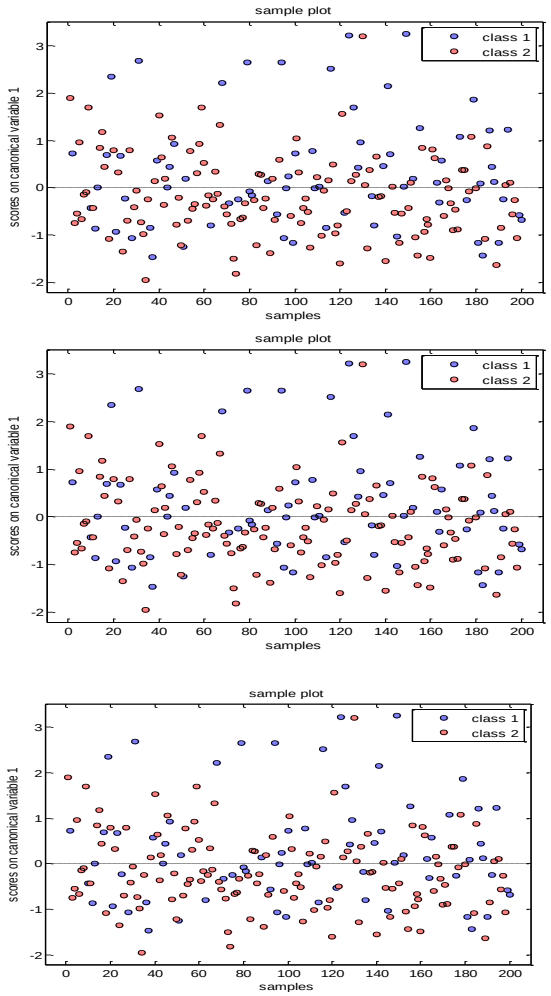


Fig. 2: Cross Validation Error Rate on Diabetes Dataset for the Different Segmentations 2, 3, 4, 5, 10, and all data as Computed by PCA Method



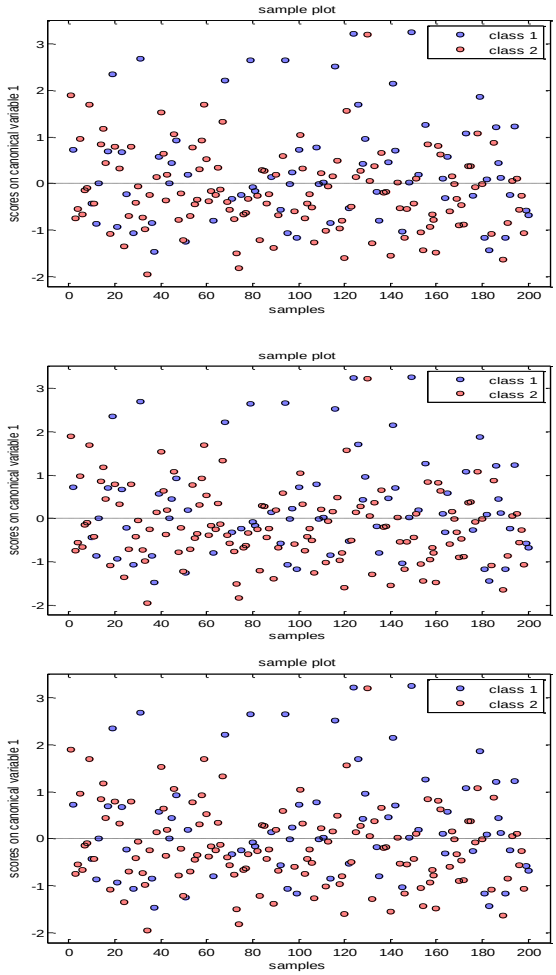


Fig. 3: The Projection of Diabetes Dataset for the Different Segmentations 2, 3, 4, 5, 10, and all Data Corresponding to Attribute 1 by PCA Method

The choice of k affects the performance of KNN method as shown in Figure (1) and the cross validation is used to choose the optimal value of k. Figure (2) shows that error rate at each attribute using cross validation with different segmentation.

V. CONCLUSION

In this work different classification methods are discussed and demonstrated by applying different datasets. By analyzing the experimental results it is observed that the PCA method has better results than other two algorithms. Also, it is found that the PCA a useful approach when dealing with large amount of data. For dataset has large number of attributes it is preferred to use the PCA method. The CART and KNN algorithms have poor performance for datasets have large number of attributes. At last, we can say that no one algorithm is the best for all types of dataset. The overfitting is that the method doesn't fit the test error as it fits the training error. The cross validation is a way used to predict the fit of the method. So, cross validation is used to estimate the expected level of fit of a method independent of the training set. The optimal value of k in the KNN algorithm is obtained by means of cross validation procedures. For all the classification methods the time to classify the instance is related to the number of instances and the number of attributes

REFERENCES

[1] A. K. Sharma and S. Sahni, "A Comparative Study of Classification Algorithms for Spam Data Analysis", International Journal on Computer Science and Engineering (IJCSE), Vol. 3, May 2011, pp. 1890-1895.

[2] T. C. Sharma and M. Jain, "WEKA Approach for Comparative Study of Classification Algorithm", The International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 4, April 2013, pp.1925-1931.

[3] S. Saha and R. Chaki, "Application of Data Mining in Protein Sequence Classification", The International Journal of Database Management Systems (IJDMS), Vol. 4, No. 5, October 2012, pp. 103-118.

[4] A. Kuwadekar and J. Neville, "Rational Active Learning for Joint Collective Classification Models", The Proceedings of the 28th International Conference on Machine Learning, Bellevue, USA, 2011, pp. 1-8.

[5] V. Korde and C. N. Mahender, "Text Classification and Classifiers: A Survey", The International Journal of Artificial Intelligence and Applications (IJAIA), Vol. 3, No. 2, March 2012, pp. 85-99.

[6] Sonam Narwal and Kamaldeep Mintwal, "Comparison the Various Clustering and Classification Algorithms of WEKA Tools", The International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3 Issue 12, December 2013, pp. 866-878.

[7] S. Neelamegam and E. Ramaraj, "Classification Algorithm in Data Mining: An Overview", The International Journal of p2p Network Trends and Technology (IJPTT), Vol. 4, Issue 8, September 2013, pp. 369-374.

[8] Nikhil N. Solvithal and R. B. Kulkarni, "Evaluating Performance of Data Mining Classification Algorithm in Weka", International Journal of Application or Inovation in Engineering and Management (IJAIEM), Vol. 2, Issue 10, October 2013, pp. 273-281.

[9] M. Tiwari and M. Mishra, "Accuracy Estimation of Classification Algorithms with DEMP Model", The International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 11, November 2013, pp. 4222-4225.

[10] S. Vijayarani and M. Muthulakshmi, "Comparative Study on Classification Meta Algorithms", The International Journal of Innovative Research in computer and Communication Engineering, Vol. 1, Issue 8, October 2013, pp. 1768-1774.

[11] RohilArora and Suman, "Comparative Analysis of Classification algorithms on Different datasets Using WEKA", The International and journal of Computer Application, Vol. 54, No. 13, September 2013, pp. 21-25.

[12] S. Vijayarani and M. Muthulakshmi, "Comparative analysis of Bayes and lazy Classification Algorithms", The International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 8, August 2013, pp. 3118-3124.

[13] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, Philip S. Yu, Zhi-Hua Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 Algorithms in Data Mining", KnowlInfSyst, Vol. 14, 2008, pp.1–37.

[14] J. Gou, L. Du, Y. Zhang, and T. Xiong, "A New Distance-Weighted K-Nearest Neighbor Classifier", The Journal of Information &Computational Science, Vol. 9, June 2012, pp. 1429-1436.

[15] T. Neo and D. Ventura, "A Direct Boosting Algorithm for the K-Nearest Neighbor classifier Via Local Wraping of the Distance Metric", Pattern recognition letters, Vol. 33, 2012, pp. 92-102.

[16] F. Fernandez and P. Isasi, "Local Feature Weighting in Nearest Prototype Classification", The IEEE Transactions on Neural Networks, Vol. 19, No. 1, January 2008, pp.40-53.

[17] O. Sutton, "Introduction of K Nearest Neighbour Classification and Condensed Nearest Neighbour data Reduction", Februray 2012, pp.1-10.

[18] L. Gordon, "Using Classification and Regression Trees (CART) is SAS Enterprise Miner for Applications in Public Health", Data Mining and Text Analysis, SAS Global Forum, 2013, pp. 1-8.

[19] W. Y. Loh, "Classification and Regression Trees", WIRES data Mining and Knowledge Discovery, John Wily&Sons, Vol. 1, January/February, 2011, pp.14-23.

[20] T. N. Phyu, "Survey of Classification Techniques in Data Mining", Proceeding of the International Multi-Conference of Engineers and Computer Scientists, IMECS 2009, March 18-20, 2009-Hong Kong, pp. 1-5.

[21] S. Arlot and A. Celisse, " A Survey of Cross Validation Procedures for Model Selection", Statistics Survey, Vol. 4, ISSN: 1935-7516, 2010, pp. 40-79.

[22] Wikipedia the free Encyclopedia "http://en.wikipedia.org/Wiki/Cross-validation-(statistics), Downloaded 2015.

[23] H. S. Yazdi and N. Salehi, "Multi Branch Decision Tree: A New splitting Criterion", International Journal of Advanced Science and Technology", Vol.45, August 2014, pp. 91-106.

[24] Downloaded from website www.sebastianraschka.com/Articles/2014_pca_step_by_step.html , 2016.

[25] M. AkhilJabbar, B. L. Deekshatulu, and Priti Chandra, "Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm", The International Conference on Computational Intelligence:Modling Techniques and Applications, Elsevier, 2013, Vol. 10, pp. 85-94