International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Comparative Study of Classification Algorithms in Chronic Kidney Disease

Pratibha Devishri S, Ragin O R, Anisha G S

Abstract*: Chronic Kidney Disease is a very dangerous health problem that has been spreading globally due to alterations in life style such as food habits, changes in the atmosphere, etc. So it is essential to decide on any remedy to avoid and to predict the disease in early stage which helps to avoid wastage of life. We show that feature selection approach is well suited for chronic kidney disease prediction. Principal Component Analysis is one of the feature selection techniques that filters out less important attributes; it also picks attributes of importance from the dataset. We also compare different data classification approaches in terms of how accurately they predict chronic kidney disease. We examine Decision stump, Rep tree, IBK, K-star, SGD and SMO classifiers using performance measures like Kappa statistics, Receiver Operating Characteristic, Mean Absolute Error and Root mean squared Error using WEKA. Accuracy measures used to compare classifiers are Recall , F-measure and Precision by implementing on WEKA. WEKA-a software for data mining, that uses collection of algorithm for data mining. It is possible to apply these algorithms directly to the data or call them from java code. Results obtained show better accuracy measures for Decision stump and Rep tree where the mean absolute error were less with error rate of 0.010 and 0.012 respectively.

Index Terms: Chronic Kidney Disease, Principal Component Analysis , Decision stump , Rep tree, IBK , K-star ,SGD , SMO Recall ,F-measure, Precision

I. INTRODUCTION

Data mining tries to assort and explore datasets and construct models from vast data stores to reveal formerly unfamiliar patterns**.** Researching healthcare data with ancient tools is not feasible due to the complexity and unclean nature of the information. To facilitate and make accurate the decision-making method, data mining techniques are implemented to change input data helpful information and extract valuable results and patterns and trends out of vast amounts of data.

Revised Manuscript Received on 30 May 2019.
***** Correspondence Author
Pratibha Devishri S*, PG Student, Master of Computer Application, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi - 682 024, Kerala.
Ragin O R, PG Student, Master of ComputerApplication, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi - 682 024, Kerala.
Anisha G S , Faculty Associate, Dept. of CS & IT, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Edappally North P.O. Kochi-682024, Kerala.. Qualification – MCA , M.Phil (CS).Area of Interest - Networks and Data mining

Data mining has found applications in several areas. It is used as a necessary tool in health care management. All agents in an exceedingly health care trade will considerably profit data mining applications.One of the semi-permanent clinical conditions in which the kidneys do not function effectively is chronic kidney disease (CKD).Sometimes, its symptoms do not manifest until advanced stages. Early detection is possible via blood and excreta tests. Complications include high vital sign, anemia ,bone weakening, nerve damage. Moreover, sickness will increase risk of getting heart and vessel disease. These issues might happen slowly over a protracted amount of your time. Chronic kidney could also be caused by polygenic disease, high vital sign and different disorders.

The potential value of early detection and treatment is great. Untreated of poorly managed, the disease results in renal failure, which needs chemical analysis or organ transplant to save the patient's life.

Our primary goal is to make a comparison study to reveal best classification algorithm techniques with their accuracy measures for chronic kidney disease. Various classification algorithm that are used are Decision stump, Rep tree, IBK, K-star, SGD and SMO classifiers using performance measures like Kappa statistics, Receiver Operating Characteristic, Mean Absolute Error and Root mean squared Error using WEKA tool. We are also undertake a comparison of the classifiers on various accuracy measures like Recall , F-measure and Precision,; a WEKA implementation was used for this.

II. RELATED WORK

Deepti Sisodiaa et al . , discuss diabetes prediction using three classification algorithms. They present a model which can predict the possibility of diabetes developing with high accuracy. The algorithm that they used are SVM , Naive Bayes and Decision Tree,. Their experimental results prove that, Naïve Bayes classification algorithm outperformed well as compared to other two algorithms [1]. According to Maryam Soltanpour et al . , chronic kidney disease is predicted using data mining techniques. Their main objective is on applying various classification algorithm to a dataset for analysis of chronic kidney disease.Various algorithm that are used as follows: Naive Bayesian ,Linear Regressing, ,Decision Tree , Neural Network and Super Vector Machine. Also, calculated and compared the measurements of performance of various techniques , previous to and subsequent to the application of feature selection [2]. Milandeep Arora and Er. Ajay Sharma examine Detection of Chronic Kidney Disease - they use WEKA to analyze Medical Datasets.

Using the mining tool weka ,their main aim is to predict chronic kidney disease and algorithms used are J48, Naïve Bayes and SMO. These algorithms are compared with one another with accurately classified instances, absolute error mean and kappa statistic as criteria. Experimental results shows that J48 is the best performing algorithm[3]. Krupa Joel Chabathula et al., undertake a comparative analysis of the approach to intrusion detection using machine learning algorithms based on feature selection method.They primarily compare approaches to intrusion detection –with or without PCA using K-Nearest Neighbors (KNN), Random Forest Tree classification, J48 Tree, Support Vector Machines (SVM), Naive Bayes probabilistic classifier , Nearest Neighbors generalized Exemplars, Voting Features Interval and Adaboost classification machine learning algorithms. Experimental results clearly show that Tree classification achieves superior results compared to other algorithms[4]. Parul Sinha and Poonam Sinha look at relative study of prediction of chronic kidney disease using SVM and KNN.They evaluate the performance of KNN and SVM choosing as criteria run time ,precision and accuracy for CKD prediction. Experimentally, it was found that KNN classifier performs better than SVM[5]. N.Radha and S.Ramya discuss performance analysis of algorithms for the prediction of chronic kidney disease in machine learning.They try to reduce and improve accuracy and diagnosis time respectively through classification algorithms. The experimental outcomes performed on different algorithms such as Decision Tree, Support Vector Machine, Naive Bayes and K-Nearest Neighbour. The experimental result found that the K-Nearest Neighbour algorithm gives better result than the other classification algorithms[6]. Sushilkumar Kalmegh discusses the classification of Indian News using analysis of weka data mining algorithm Random Tree ,Simple Cart and REPTree. The aim of their work is to compare in terms of performance, RandomTree, REPTree and Simple Cart classification algorithms. They find that on Indian news, Random Tree is more efficient and accurate than Simple Cart and Rep tree [7]. Anu Chaudhary and Puneet Garg study patient monitoring system for the detection and diagnosis of a disease.Their intnent is to predict heart disease, kidney failure using K-mean and A-priori algorithms. They conclude that finding the stage of heart disease and kidney failure is an easy and efficient way [8]. W. John Wilbur and Won Kim discuss about MeSH Prediction and Stochastic Gradient Descent. In this paper, they compare the stopping predictions with simply stopping at a fixed number of iterations. They chose three algorithms for comparison: PROBE, SVMperf, and Liblinear and compare them with SGD.The experimental results shows that SGD-SVM predictions for all MeSH terms and utilizes the Pool Adjacent Violators (PAV) algorithm to change these predictions to probabilities [9]. Guneet Kaur and Ajay Sharma discusses predict CKD using Hadoop algorithms for data mining. They try to determine CKD dataset using classification algorithms like K-Nearest Neighbor(KNN), Support Vector Machine(SVM), Naive Bayes and dimensionality reduction algorithms like Independent Component Analysis(ICA) and Principle Component Analysis(PCA)in MATLAB by obtaining Hadoop to examine a single patient for the disease and to disclose the elapsed time[10]. Kerina Blessmore Chimwayi et al., discuss prediction of the risk level of CKD. They use Hierarchical Clustering Algorithm and Neuro-Fuzzy. In this paper, they applied Neuro-fuzzy algorithm to conclude the risk of CKD in patients using selected features. The outcomes of the predictions are grouped to identify the percentage of patients with high chance of having kidney disease who have a higher likelihood of being diabetic[11]. P. Suguna and Dr. S. Prema discusses J48 Decision Tree and Novel Genetic Algorithm Framework for Chronic Kidney Disorder (CKD). This study focuses on the pre- processing of noisy data, applied algorithms for machine learning,improving the detection of glomerular filtration rate (GFR) , predicting the problem for patients with CKD . The experimental results shows that Genetic algorithm appear to be best for the problem[12]. Sirage Zeynu and Shruti Patil discusses Survey on Prediction of CKD Using DataMining Classification Techniques and Dimensionality Reduction Method. The main focus is to determine and predict Chronic Kidney Disease by detecting the unseen pattern by using dimensionality reduction algorithm and classification techniques like decision tree , artificial neural network (ANN), K-nearest neighbor (KNN). Experimental results found that data mining classification techniques and dimensionality reduction algorithm have been used to determine, recognize and predict chronic kidney and other diseases[13]. Dr. S. Vijayarani and Mr.S.Dhayanand discuss data mining algorithms for the prediction of kidney disease.The main aim is to predict and find the best algorithm for classification using Support Vector Machine and Naïve Bayes based on accuracy and execution time. SVM has higher performance as comparing to Naïve Bayes,experimentally[14]. K.R.Lakshmi et.al. discuss performance comparison of three techniques of data mining to predict the survival of kidney dialysis. The main objective of this work is to establish forecast model for the survival of kidney dialysis and to find the best algorithm.In this paper, they used three classification algorithm - Logical Regression , Decision tree and Artificial Neural Networks . The final outcome shows that ANN performed well than other algorithms based on accuracy and performance [15].

III. METHODOLOGY

A. Dataset

To classify data, analyze data and estimate the prediction accuracy of different data mining algorithms , we use WEKA as a data mining tool. We analyze the accuracy of different algorithms and identify the best suited one in terms of classification accuracy. We suggest techniques best suited to the classification and prediction of chronic kidney disease from medical datasets. We use datasets released by the UC Irvine Machine Learning Repository. Using WEKA, we implement different classification algorithms on this dataset. The dataset used for principal component analysis consists of 400 instances out of which were present 250 instance of CKD detected and 150 of NOTCKD detected. The dataset consists entities with a total of 24 attributes each out of which 11 were numeric and 14 nominal value attributes. We will be implementing Principal Component Analysis to select the relevant data set. After preprocessing the data set, it consists of 14 attributes and one class attribute. Using this dataset, we spilt these dataset into two different set-training set and test set.

There are 400 instances which spilt into training set and test set without overlapping that means the set will not contain duplicate items. In training set, it contains 70% of data that is 280 instances and the rest of the 30% of data are considered as cross validation and test data. After training the data, test the data to the training set. At the end, the result file shows the predicted results. These results are analyzed experimentally with WEKA tool.

Sl No	Attribute Used	Type Of Attribute	Attribute Description
1	Age	Numerical	Age
2	Bp	Numerical	Blood pressure
3	Sg	Nominal	Specific gravity
4	Al	Nominal	Albumin
5	su	Nominal	Sugar
6	rbc	Nominal	Red blood cells
7	pc	Nominal	Pus cell
8	pcc	Nominal	Pus cell clumps
9	ba	Nominal	Bacteria
10	bgr	Numerical	Blood glucose random
11	bu	Numerical	Blood urea
12	se	Numerical	Serum creatinine
13	sod	Numerical	Sodium
14	pot	Numerical	Potassium
15	hemo	Numerical	Haemoglobin
16	pcv	Numerical	Packed cell volume
17	we	Numerical	White blood cell count
18	re	Numerical	Rbc count
19	htn	Nominal	Hypertension values
20	dm	Nominal	Diabetes mellitus
21	cad	Nominal	Coronary artery disease
22	appet	Nominal	Appetite
23	pe	Nominal	Pedal edema
24	ane	Nominal	Anaemia
25	class	Nominal	Class

Fig. 1 Chronic Kidney Disease Dataset

B. Model Diagram

Proposed model states that Feature selection method is applied on original data and it is preprocessed to eliminate attributes which is less important. 70 % of these data are trained and 30 % of remaining data are used for testing. Six different classification algorithms including Decision stump, Rep tree, IBK , K-star ,SGD and SMO are applied on the trained data and data set obtained via Stratified cross-validation and the supplied test data are compared based on how the classification algorithms perform and evaluated in terms of instances that are correctly classified, Kappa statistics , Receiver Operating Characteristic, Mean Absolute Error and Root mean squared Error .These are then compared on various accuracy measures like Precision, Recall and F-measure .

We summarize the proposed procedure in Figure-2 .

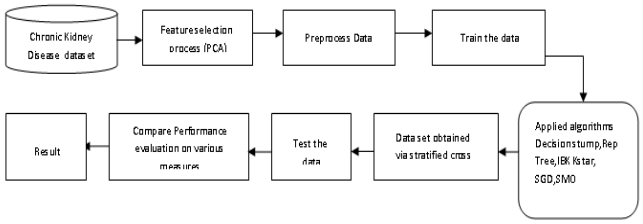

Fig. 2 Proposed Model Diagram

Principal Component Analysis: Feature Selection Method

Principle Component Analysis (often referred to as PCA) could be a common technique used in knowledge science to extract features. It calculates the covariance matrix 's eigen vectors with optimal Eigen values; it then proceeds to project the information into a new subspace of equal or lesser dimensionality. Thus, PCA transforms a matrix of n' options into an all new dataset of options but n'.

Indeed, let M be the matrix, 'm' be the length of the vector and 'n' the number of terms . Then M can be represented as follows:

Each column stands for a vector in equation 1.

$$M_{m \times n'} = [M_1, M_2, M_3, ..., M_{n'}] \quad (1)$$

For each vector, we evaluate the average mean ().

$$\mu_1 = \frac{1}{n'} \sum_{i=1}^{n} M_{ij} \quad (2)$$

We then proceed to estimation of deviations:

$$\Phi t = M_i - \mu_1 \quad (3)$$

Now, based on the number of classes, we assemble the co-variance matrix

$$N_{m \times n'} = \frac{1}{(n'-1)} \sum_{i=1}^{n'} \emptyset t_i \emptyset_i' \frac{1}{(n'-1)} \sum_{i=1}^{n'} (X_i - \mu_1)(X_i - \mu_1)' \quad (4)$$

In the above, Φ' is inverse of the matrix Φt.

The reduced subspace is then evaluated:

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{m} \lambda_i} \geq S' \quad (5)$$

Here, 'S' ' is the ratio between the reduced subspace variation and the total high- dimensional space variation. According to equation 6, the data represented by the principal components in the reduced dimension P subspace.

$$Y_i = U_1' \Phi_i = U_1'(X_i - \mu_1) \quad (6)$$

Where D_1 = weight vector and 'i' = number of training.[4]

C. Classification Algorithms

1) Decision Stump

Decision stump is a model for machine learning. It consists of a decision tree at a single level. This tree is a decision tree directly linked to each leaf with a root node. Decision stump creates a production determined by the value assumed by exactly one input feature. They are also called 1 rule from time to time. As the input function changes, there may be several variations. Decision stump are frequently used as components in ensemble techniques for machine learning.Eg, they feature in bagging and boosting.

=== Confusion Matrix of Decision Stump (Stratified cross-validation) ===

```
  a   b   <-- classified as
171   2 |   a = ckd
  0 107 |   b = notckd
```

2) Rep Tree

RepTree employs the logic of regression tree and builds numerous trees using distant iterations. After that it chooses the best tree from all produced and treat it as the representative. The tree is then pruned; the measure utilized is the average square error of the predictions resulting from the tree. The algorithm of REP Tree is a quick decision tree learner. It uses info gain / variance to construct a decision / regression tree and dock it with reduced-error pruning (with backfitting). For numeric attributes, the values undergo sorting only once. To deal with missing values, we divide the corresponding instances into items.

=== Confusion Matrix of Rep tree (Stratified cross-validation)===

```
  a   b   <-- classified as
171   2 |   a = ckd
  0 107 |   b = notckd
```

3) IBK

IBK is the implementation of an instance- based learning or lazy learning neighbor in K- nearest where function is almost accurate and until classification ,all the calculations are delayed. Comparing to other algorithm, IBK is the most easiest machine learning algorithm.

=== Confusion Matrix of IBK (Stratified cross-validation)===

```
   a   b   <-- classified as
 168   5 |   a = ckd
   0 107 |   b = notckd
```

4) K-star

K-star is a classifier based on instances, which is a test instance class. It has as foundation, the class of those training instances similar to it, as it is driven by some similarity function. It uses distance operation based on entropy; this makes it stand apart from distant learners based on instances.

=== Confusion Matrix of Kstar (Stratified cross-validation) ===

```
   a   b   <-- classified as
 146  27 |   a = ckd
   0 107 |   b = notckd
```

5) SGD

Stochastic gradient descent (frequently shortened to SGD), additionally referred to as progressive gradient descent, is an unvaried methodology for optimizing a differentiable, objective perform a random approximation of gradient descent improvement. Gradient descent can be utilized to train different kinds of regression and classification models. It is an iterative process and therefore is appropriate for map reduce process. Then pseudo code for the SGD algorithm without regularization is as follows.[9]

SGD without regularization

- Input: training data. $\{(x_i y_1)\}_{i=1}^{n}$; learning rate λ>0.
- Initialize: $w_0 = 0$; $t1 = 0$
  1. Randomly sample j, $1 \le j \le N$
  2. If $y_i (w_t \cdot x_j) < 1$ set $w_{t+1} = w_t + \lambda y_j x_j$ and $t1 = t1 + 1$
  3. If stopping criterion satisfied return w_t else return to 1.[9]

=== Confusion Matrix of SGD (Stratified cross-validation) ===

```
   a   b   <-- classified as
 170   3 |   a = ckd
   0 107 |   b = notckd
```

6) SMO

Sequential minimal optimization (SMO) is a method for eliminating the quadratic programming (QP) drawback – this arises at all stages of the training of support vector machines. It is iterative and finds use in solving optimization problems. SMO splits this drawback into a sequence of smallest tractable sub-problems which are then attacked analytically.

The result of the SMO algorithm is as follows:

- Locate a Lagrange α_1 multiplier that infringes the Karush-Kuhn-Tucker optimization conditions.
- Choose a second α_2 multiplier and optimize the pair (α_1, α_2).
- Repeat the above steps until merging

=== Confusion Matrix of SMO(Stratified cross-validation)===

```
   a   b   <-- classified as
 167   6 |   a = ckd
   0 107 |   b = notckd
```

IV. Results and Discussion

Six different classification algorithms were applied on the trained data and data set obtained via Stratified cross-validation and the supplied test data are compared based on the performance of classification algorithms and evaluated based on instances that were correctly classified, Kappa statistics, Receiver Operating Characteristic, Mean Absolute Error and Root mean squared Error. This was done by choosing Recall, F-measure , Precision etc as criteria. . In Table 1-3, results are shown, by comparing and examining the results it is understood that Decision stump and Rep tree are giving better results in terms of Recall, F-measure, Precision, Kappa statistics with less error rate of 0.010 and 0.012 respectively.

Algorithm	Precision	Recall	F-measure	ROC
Decision Stump	0.993	0.993	0.993	0.992
Rep tree	0.993	0.993	0.993	0.991
IBK	0.983	0.982	0.982	0.986
K- star	0.923	0.904	0.905	0.998
SGD	0.990	0.989	0.989	0.991
SMO	0.980	0.979	0.979	0.983

Table I: Performance of algorithms for classification. Criterion: Accuracy Measures.

Algorithm	Root mean squared error	Mean absolute error	Kappa Statistic
Decision Stump	0.084	0.010	0.984
Rep tree	0.084	0.012	0.984
IBK	0.133	0.021	0.962
K- star	0.292	0.101	0.805
SGD	0.103	0.010	0.977
SMO	0.146	0.021	0.955

Table II: Performance of algorithms for classification. Criterion: error rate and kappa statistics

Classification Algorithms	Correctly Classified Instances	Incorrectly Classified Instances
Decision Stump	278	2
Rep tree	278	2
IBK	275	5

K- star	253	27
SGD	277	3
SMO	274	6

Table III: Performance. Criterion: Classified Instances

V. CONCLUSION AND FUTURE WORK

It is essential to predict chronic kidney disease accurately as it is stated as a deadly disease. Chronic Kidney Disease is predicted using six classifiers. We compared the outputs with accuracy achieved as the criterion. These algorithms are compared using instances that they classified correctly, Kappa statistics , Receiver Operating Characteristic, Mean Absolute Error, Root mean squared Error , F-measure ,Recall, and Precision. Experimental results show that, out of six classifiers, Decision Stump and Rep tree performed better accuracy measure than other algorithms. K-Star shows more accuracy measure in ROC. Decision stump and Rep tree algorithms are giving better results with less error rate. We expect that the designed system backed up with the classification algorithms used may be deployed to anticipate or analyze data pertaining to other disorders.

REFERENCES

1. Deepti Sisodiaa , Dilip Singh Sisodia, Prediction of Diabetes using three classification algorithms, International Conference on Computational Intelligence and Data Science(ICCIDS 2018), Procedia Computer Science 132 (2018) 1578–1585
2. Maryam Soltanpour Gharibdousti, Kamran Azimi, Saraswathi Hathikal, Dae H Won , Prediction of Chronic Kidney Disease using Data mining techniques , 2017 Industrial and Systems Engineering Conference
3. Er. Ajay Sharma , Milandeep Arora , Chronic Kidney Disease Detection by Analyzing Medical Datasets in Weka , International Journal of Computer Application (2250-1797) Volume 6– No.4, July- August 2016
4. Krupa Joel Chabathula, Jaidhar C.D , Ajay Kumara M.A Comparative Study of Principal Component Analysis Based Intrusion Detection Approach Using Machine Learning Algorithms , 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN)
5. Parul Sinha ,Poonam Sinha , Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM , International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV4IS120622 , Vol. 4 Issue 12, December-2015
6. N.Radha , S.Ramya , Performance Analysis of Machine Learning Algorithms for Predicting Chronic Kidney Disease ,International Journal of Computer Sciences &Engineering(ICSE), Volume-3, Issue-8 E-ISSN: 2347-2693[2015]
7. Sushilkumar Kalmegh , Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and RandomTree for Classification of Indian News, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 2, February 2015.
8. Anu Chaudhary, Puneet Garg , Detecting and Diagnosing a Disease by Patient Monitoring System , International Journal of Mechanical Engineering and Information Technology (IJMEIT), Vol. 2 Issue 6 ,June 2014,493-499 [2014]
9. W. John Wilbur, Won Kim , Stochastic Gradient Descent and the Prediction of MeSH for PubMed Records, AMIA Annual Symposium Proceedings, 2014 Nov 14;2014:1198-207.
10. Guneet Kaur , Ajay Sharma , Predict Chronic Kidney Disease Using Data Mining Algorithms In Hadoop , International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-5, Issue-6, Jun-2017
11. Kerina Blessmore Chimwayi, Noorie Haris, Ronnie D. Caytiles*and N. Ch. S. N Iyengar**, Risk Level Prediction of Chronic Kidney Disease Using Neuro-Fuzzy and Hierarchical Clustering Algorithm (s)
12. P. Suguna, Dr. S. Prema ,J48 Decision Tree and Novel Genetic Algorithm Framework for Chronic Kidney Disorder (CKD) , International Journal of Advanced Research in Science,Engineering and Technology,Vol. 4, Issue 11 , November 2017
13. Sirage Zeynu and Shruti Patil ,Survey on Prediction of Chronic Kidney Disease Using DataMining Classification Techniques and Feature Selection, International Journal of Pure and Applied Mathematics,Volume 118 No. 8 2018, 149-156
14. Dr. S. Vijayarani and Mr.S.Dhayanand , Data Mining Classification Algorithms For Kidney Disease Prediction , International Journal on Cybernetics & Informatics (IJCI) Vol. 4, No. 4, August 2015
15. K.R.Lakshmi, Y.Nagesh and M.VeeraKrishna , Performance Comparison Of Three Data Mining Techniques For Predicting Kidney Dialysis Survivability , International Journal of Advances in Engineering & Technology, Mar. 2014.

AUTHORS PROFILE

Pratibha Devishri S PG Student, Master of Computer Application, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi - 682 024, Kerala.

Ragin O R PG Student, Master of Computer Application, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi - 682 024, Kerala.

Anisha G S , Faculty Associate, Dept. of CS & IT, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Edappally North P.O. Kochi-682024, Kerala.. Qualification – MCA , M.Phil (CS).Area of Interest - Networks and Data mining