Cluster Comput
DOI 10.1007/s10586-017-0933-1

Improved credit card churn prediction based on rough clustering and supervised learning techniques

R. Rajamohamed[1] · **J. Manokaran**[1]

Received: 16 February 2017 / Revised: 3 May 2017 / Accepted: 18 May 2017

Abstract Every process is digitized in the current society. Transfer of money from one account holder to another has become possible in seconds because of advanced technologies in information processing. Not only this in all sectors like railways, insurance, health sector, fashion technology, education sector, sales and business sectors, and advertisement sectors every firm digitized its operations. One such sector is banking where every individual based on his or her financial status will be considered for crediting loan, and credit card etc. If the credit score of the loan availing person is high banks will be ready to provide him with the loan but the availing person can opt for any one of the banks on his or her own willing. Such scenario happens in credit card churn prediction also. Hence the banks should take healthy measures to retain the existing credit card holders without any churn. Withholding existing customers of a firm plays an important role to increase the overall revenue of the firm and retains the good name of the firm in competitive market. Hence every organization takes key measures to withhold existing customers using customer management models. Because customer retention is a crucial task as it reduces the time, money and workforce needed for adding new customers to the firm. Customers retention technique in credit card churn prediction (C3P) was done using only supervised classification techniques. But it could not end with better results. So, through many proven hybrid classification techniques we can bring better accuracy in C3P. Also C3P lags in highly efficient techniques like rough set theory. Hence in this work initially we perform data processing techniques and in second stage we propose modified rough K-means algorithm used for clustering credit card holders and in next stage hold-out method divides the cluster data into testing and training clusters. At last classification is performed using various algorithms like support vector machine, random forest, decision tree, K-nearest neighbor, and Naive Bayes. Finally we evaluate the work using precision, recall (sensitivity), specification, accuracy, and misclassification error.

Keywords Credit card · C3P · Churn · K-means · Modified rough K-means · Decision tree · Support vector machine · Random forest · Naive Bayes · K-nearest neighbor

1 Introduction

Banks play a key role in the development of country in both socio and economic status. Money plays a vital role in the development of every individuals growth. Digitization of bank accounts has possible due to the development of information retrieving, storing and processing techniques. Not only all the above but growth in the expansion of each bank has become possible due to the money transaction and lending capacity of the bank. All these well known and recognized growths are due to the customer base of the particular bank. Hence retention of the existing customers plays a key role in growth of the bank [1]. Various banks offer attractive plans like savings account with nil balance, debit card, credit card, NEFT, RTGS facilities, mobile banking, internet banking, credit points based on the usage of customers, best low interest rates for various loans like housing loan, education loan, vehicle loan, mortgage loan, agricultural loan, cattle farm implementation loan etc,. Among all these functions crediting loan to a customer is a crucial task where each bank has to analyze the capacity of the customer in prior to offer that loan. In order to complete the above process

✉ J. Manokaran
manoraj3@gmail.com

1 Department of Electronics and Communication Engineering, Indra Ganesan College of Engineering, Tiruchirappalli, India

many banks have decided to incorporate a special scheme called credit card where each customer applying for credit card will be checked for his or her eligibility to avail that card. Based on the credit points many banks initiate the request for providing credit cards to the new customers. However every customer can have more than one credit card in more than one bank so there will lot of chances for a customer to churn out of particular bank[2]. Hence it is the responsibility of Bank Credit Card Account Management System (BCCAMS) to preserve the existing customers through low interest rates, and high period of return. Through highly brilliant data mining systems it is possible to design prediction system that will tell us whether the credit card holders will churn out of particular bank or not based on various available attributes collected from the past history of old customers. Data mining methods such as decision tree, naive bayes, logistic regression, random forest, artificial neural networks, inductive rule learning, and support vector machine find out the churn[1,3–8]. All these techniques are implemented not only in banking but also in medical systems, insurance, telecommunication, gaming, automobile industries, retail marketing etc.,[9–23]. Previously customers retention technique in credit card churn prediction was done using only supervised classification techniques [1–8,24–29]. But it could not end with better results. So, through many proven hybrid classification techniques we can bring better accuracy in C3P[9–17,30]. Also C3P lags in highly efficient techniques like rough set theory [31]. Hence in this work initially we collect credit card churn prediction data from UCI bench marked data set repository which is used in data processing techniques to filter missing values, any outliers, and redundancy present in it. In second stage we propose modified rough K-means algorithm used for clustering credit card holders. And in next stage hold-out method divides the cluster data into testing and training clusters. At last classification is performed using various algorithms like support vector machine , random forest, decision tree , k-nearest neighbor , and naive bayes . Finally we evaluate the work using precision, recall (sensitivity), specification, accuracy, and misclassification error.

2 Literature survey

Research on credit card churn prediction system can be implemented using many data mining techniques that are constantly being used. Such survey of available methods can be listed out briefly here. Yeh et al. in 2009 compared data mining techniques for the predictive accuracy of probability of default of credit card clients using sorting smoothing method involving six different mining techniques like k-nearest neighbor, logistic regression, naive bayes, artificial neural networks, classification trees, and discriminant analysis evaluated using lift charts. Of the above methods artificial neural networks provided better results with 0.54 validation accuracy [1]. Oyeniyi et al. applied data mining techniques to predict customer churn using Nigerian bank real life data set. With the help of WEKA tool simple k-means clustering algorithm was applied to form clusters and JRIP was used to generate rules for classification. From this they had developed patterns to identify likely to churn customers [2]. Anil Kumar et al. proposed churn prediction system using ensemble classifier, Multilayer perceptron , logistic regression , decision trees , random forest , radial basis function network and support vector machine . Business Intelligence Cup data set from the university of chile in 2004 was used. Four varieties of sampling techniques like under sampling, over sampling, a combination of under and over samplings, and synthetic minority oversampling method for balancing the customer data set to build models. They had also deployed classification and regression tree for feature selection. They identified the fourth sampling method worked very well with high accuracy [24]. A rule based churn prediction using rules extracted from support vector machine was proposed by Farquad et al. in the year 2009 and they proved better sensitivity with the hybrid model of SVM and naive bayes tree. The data set was obtained from Business Intelligence Cup 2004 [3]. In 2009 Nie et al. from china used logistic regression to build a model on credit card churn prediction using the banking data obtained from 5000 customers and proved that the system was performing very well [25]. In 2011 Nie et al. made a bank credit card churn prediction model using regression and decision tree in which they considered four parameters like customer, card, risk, and transaction activity information [4]. In 2009 Wang et al. predicted credit card churn using rough set theory and least square SVM and proved highest accuracy rate of 89.90[5]. Wang et al. in 2010 predicted the same using multi criteria churn prediction techniques and provided better results [26]. In 2003 Chiang et al. constructed goal based predictive systems using sequential patterns that was able to alarm organizations before losing valuable customers and thus it was helping a lot in decision making process [27]. In 2009 Jinbo et al. used three varieties of Ada-Boost algorithms like real, gentle, and modest Ada-Boost algorithms to prove better predictive accuracy compared to other models like support vector machine [28]. Alfred et al. in predicted customer churn using knowledge extracted from the emergent structure maps [29]. From the above reviews the hidden information we obtained is all people who had performed churn prediction in credit card bank data were using only single classifier or ensemble of classifier or hybridization of two or more classifiers but there was a lag in research on unsupervised classifiers. But when we use both supervised and unsupervised clustering in telecommunication data set we could able achieve better accuracy [9–17]. In 2013 Ying Huang et al. proposed a hybrid churn prediction system on UCI telecommunication data set in which he used modified version of k-means clustering algorithm combined with rule based inductive system(FOIL) and they proved hybrid system works very well compared to other single classifiers and tested the system on 22 bench marked data sets available [9]. In 2009 Bose et al. used five clustering algorithms like partition based, fuzzy based and self organizing map clustering methods and then boosted decision tree was used for classification and they also identified new attributes based on usage and revenue details[10]. In 2012 Huang et al. used seven classification algorithms like logistic regression, naive bayes, linear classification, decision trees, multilayer perceptron neural networks, support vector machines, and evolutionary data mining algorithms. They also constructed new features set from the existing features set and combination of seven classification algorithms and new features set proved higher efficiency [11]. In 2011 Mehdi et al. proposed hybridization of artificial neural network using multiple linear regression models and proved higher efficiency compared to traditional classification techniques [12]. In 2009 Tsai et al. used hybrid neural network based on back propagation neural networks and self-organization maps. The first technique filters noisy data and forms the clean data which is given as input to the second model. The performance is evaluated using general testing set and two fuzzy testing sets and rather than single artificial neural network , the hybrid of artificial neural network proved better accuracy [13]. V. Yeshwanth Raj et al. used evolutionary algorithms by combining genetic algorithm and decision tree to come up with rules needed for classification [14]. In 2006 Shin-Yung et al. designed a system which was giving out the churn score of each subscriber of a mobile operator and it was indicating likely to churn customers and was based on k-means combine with decision tree and artificial neural network with decision tree [15]. In 2007 Yangming et al. used k-nearest neighbor and logistic regression to create a hybrid model [16]. In 2006 Lee et al. constructed a new model called SePI which performed better than artificial neural network, decision tree, and logistic regression[17]. In 2014 Kim et al. proposed community based churn prediction model. Here they identified the initial churners who spreading churning information between their community[32]. In this paper we propose a new hybrid model a combination of rough K-means clustering algorithm and an effective five classification algorithms.

3 Churn prediction model and methods

The diagram (Fig. 1) given below pictorially depicts the working of the proposed model. In this work UCI data set on bank credit card churn prediction is used. Initially the data set consists of data ranging widely so data cleaning is applied to make it a consistent data set. Next to cleaning clustering methods like rough k-means algorithms, k-means algorithm and fuzzy c-means are used. The clusters thus got from the above process are divided into testing and training data partitions using hold-out partitioning method. The testing and training partitions are handled by supervised mining processes. The hybridized system efficiency is measured in terms of parameters like precision, recall (sensitivity), specification, accuracy, and misclassification error.

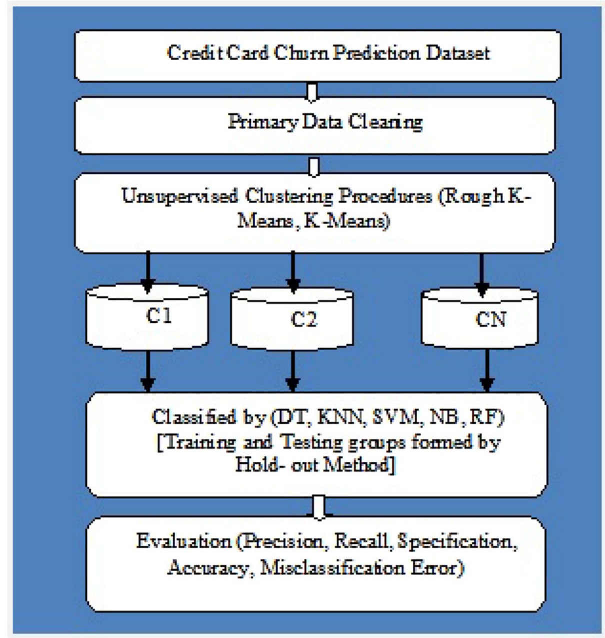

Fig. 1 Credit card churn prediction work flow

3.1 Dataset

The data set is obtained from UCI repository (a bank in Taiwan describing credit card holders of the bank) [33]. The data set totally has 30,000 member data and characterized as 23 various attributes and one churn prediction variable. Among 30,000 members 6636 members are identified as churners and 23,364 members are identified as non-churners. Each customer of the bank is given an identification number similar to uniquely identify each customer.

X1: Credit amount given to individual as well as to his or her family in dollars NT.
X2: Represents gender 1 = male, 2 = female.
X3: Education (1, 2, 3, and 4 represents graduate school, university, high school, and others)
X4: 1, 2, and 3 represents married, unmarried, and others
X5: Represents age in years.
X6–X11: Past repayment history.
X12–X17: Total due amount of the customer.
X18–X23: Amount that was paid to the bank previously.
Y:Churn variable 1 = churn, 0 = non-churn

Table 1 shows the statistical information about each attribute present in the data set.

3.2 Dataset preprocessing

Generally the dataset collected from any organization, or collected from websites or any other source if it is huge enough knowledge can be extracted to make decisions that will enhance the quality of the system in future. But the data collected will never be in good quality, it may have missing values, redundant data, noisy part, and wide range of differences between related data values. Hence it is the responsibility of miners to clean the data in the first step. Such cleaning process is called dataset preprocessing. In this work data set has wide range of values so the attribute values are normalized in the range of zero to one using min-max normalization. Also X6–X11 have some negative values which can be converted to other numeric values which are not present in the data set.

3.3 Clustering algorithm

Major data mining models use clustering as a technique in audio, video, and image processing. It aims to group related member objects with similar properties. Hence all the members within a cluster are highly similar and members between the clusters are highly dissimilar. Because of clustering a hybrid mining model can be formulated which proves better accuracy. Supervised classification techniques can be applied on formed clusters to enhance accuracy.

3.3.1 K-means

K-means is a partition based clustering algorithm formulated by Stuart Lloyd in 1957. Here K partitions which should be less than N are formed where N represents input tuples. The following properties should be hold true while using k-means clustering algorithm.

- **Property 1**: Every tuple should be a part of at least one cluster
- **Property 2**: Every cluster should hold at least one tuple

Thus the data set which contains N tuples are input into the system resulting in K number of clusters. The following steps are carried to apply k-means clustering explained in Algorithm 1.

$$D_i = \sum_{C=1}^{K} \sqrt{\left(D_{i,C} - C_C\right)^2} \tag{1}$$

Algorithm 1: K-means algorithm

1. Fix with the value K and it should be less than the value of N
2. Randomly K values are chosen as initial cluster centers
3. Next distance is measured from each cluster member to randomly chosen cluster centers using Eq. (1).
4. The cluster node becomes the cluster member of the cluster center which is at minimum distance to that node
5. Mean value of all the members of a cluster are computed which is considered as new cluster center
6. Now the steps 3, 4, and 5 are repeated until the cluster center becomes stable

3.3.2 Fuzzy C-means

Fuzzy C means is a fuzzy based clustering algorithm formulated by Dunn in 1973. Here K partitions which should be less than N are formed where N represents input tuples. The following properties should be hold true while using fuzzy c-means clustering algorithm.The following steps are carried to apply fuzzy c-means clustering explained in Algorithm 2.

- **Property 1**: Every tuple have K cluster Membership matrix
- **Property 2**: Addition of K cluster Membership matrix is equal to 1

$$\mu_{i,j} = \frac{1}{\sum_{k=1}^{c} \frac{(d_{i,j})}{(d_{i,k})}^{\frac{2}{m-1}}} \tag{2}$$

$$V_j = \frac{\sum_{i=1}^{n} (\mu_{i,j})^2 x_i}{\sum_{i=1}^{n} (\mu_{i,j})} \tag{3}$$

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{i,j})^m \left\| x_i - v_j \right\| \tag{4}$$

Algorithm 2: Fuzzy C-means algorithm

1. Fix with the value K and it should be less than the value of N
2. Randomly K values are chosen as initial cluster centers
3. Next fuzzy membership function is measured from each cluster member to randomly chosen cluster centers using Eq. (2).
4. The cluster node becomes the cluster member of the cluster center which is at maximum membership to that node
5. New cluster centers are computed using Eq. (3).
6. Now the steps 3, 4, and 5 are repeated until the objective function (Eq. (4).) is reached

3.3.3 Rough K-means

In 2004 Lingras et al. formulated a new clustering algorithm called rough K-means and applied it in web mining[34]. In 2005 Georg Peters analyzed rough k-means algorithm

Table 1 Statistical information about each attribute present C3P

S.No	Attributes	Distinct.count	Min	Max	Mean	SD	Nominal values
1	LimitBal	81	10,000	1,000,000	167,484.323	129,747.662	–
2	Sex	2	–	–	–	–	1-11888,2-18112
3	Education	7	–	–	–	–	0-14, 1-10,585, 2-14,030, 3-4918, 4-123, 5-280, 6-51
4	Marriage	4	–	–	–	–	0-54, 1-13,659, 2-15,964, 3-323
5	Age	56	21	79	35.486	9.218	–
6	Pay0	11	−2	8	−0.017	1.124	–
7	Pay1	11	−2	8	−0.134	1.197	–
8	Pay2	11	−2	8	−0.166	1.197	–
9	Pay3	11	−2	8	−0.221	1.169	–
10	Pay4	11	−2	8	−0.0266	1.133	–
11	Pay5	11	−2	8	−0.291	1.150	–
12	BillAmt1	22,723	−165,580	964,511	51,223.330	73,635.861	–
13	BillAmt2	22,346	−69,777	983,931	49,179.075	71,173.769	–
14	BillAmt3	22,026	−157,264	1,664,089	47,013.155	69,349.38	–
15	BillAmt4	21,548	−170,000	891,586	43,262.949	64,332.856	–
16	BillAmt5	21,010	−81,334	927,171	40,311.401	60,797.156	–
17	BillAmt6	20,604	−339,603	961,664	38,871.76	59,554.108	–
18	PayAmt1	7943	0	873,552	5663.581	16,563.28	–
19	PayAmt2	7849	0	1,684,259	5921.164	23,040.87	–
20	PayAmt3	7518	0	896,040	5225.682	17,606.961	–
21	PayAmt4	6937	0	621,000	4826.077	15,666.16	–
22	PayAmt5	6897	0	426,529	4799.388	15,278.306	–
23	PayAmt6	6939	0	528,666	5215.503	17,777.406	–
24	Churn	2	–	–	–	–	1-6636, 0-23,364

and formulated a new enhanced rough k-means algorithm and applied it gene expression microarray, synthetic, and forest data [31]. Parvesh Kumar applied k-means, rough k-means, and PAM clustering algorithms on cancer data sets and comparatively proved rough k-means algorithm worked with higher efficiency [35]. In 2002 Lingras et al. submitted a comparative study on conventional k-means and rough k-means algorithms in web usage mining process[36]. In 2007 Darshit Parmar framed a new algorithm for categorical data clustering, coined as MinMin-Roughness (MMR), it works on the basis of rough set theory (RST),and handles the uncertainty during clustering process [37]. Amin et al. proposed customer churn prediction using rough set theory. Initially the data set consists of data ranging widely so data cleaning is applied to make it a consistent data set. They also constructed new features set from the existing features set using information gain, then testing and training data partitions using hold-out partitioning method and they are using RSES tool kit for generating rule. Finally the test data is classified using Genetic, LEM, Covering, Exhaustive algorithm [38,39]. K-means algorithm generally uses the properties of both k-means, fuzzy concepts and rough set theory .The following steps are carried to apply rough K-means clustering explained in Algorithm 3. Rough K-means algorithm works based on the following properties:

- Each cluster has two levels lower approximation and upper approximation
- Each cluster in its lower approximation should possess at least one data member at the beginning
- Members definitely in lower approximation becomes the definite member of that cluster
- The members who are in upper approximation having fuzziness in more than one cluster can be a member of all those clusters

$$CC = wl \times \sum_{j=1}^{NA} \sum_{i=1}^{N} \frac{D_i}{N} + wl \times \sum_{j=1}^{NA} \sum_{i=1}^{N} \frac{D_i}{N} \tag{5}$$

3.4 Classification algorithm

The main aim of classification is prediction and decision making. Classifiers can be generally applied on any mining datasets intended to extract some knowledge from the process. It can handle audio, video, image, text, and numeric data sets and builds a predictive model. Based on the construction of the training and testing data sets and on the properties of the classifier a class label identifies the nature of the mining process. It can be binary level, ternary level, and multilevel classification. Here we have discussed five major classifiers that are used extensively in this research.

Algorithm 3: Rough K-means algorithm

1. Fix the value of K (cluster center) which should be less than samples N
2. Randomly divide the samples N into K number of clusters
3. Find the cluster center (CC). Here wl = 0.3, and wc = 0.7, and all the attribute values of cluster are divided individually by the total number of tuples that are present in a particular cluster and thus obtained values are added together to form a sum. The sum formed from above steps are multiplied individually with wl and wc values which represent lower and upper values when added together forms cluster center using Eq. (5). Here NA represent the number of attribute
4. Find the Euclidean distance between each data point to each cluster center and the minimum is considered as first lower approximation using Eq. (1). and it's consider as new cluster center.remaining element is consider as upper bound
5. Find the Euclidean distance between upper bound to new cluster center using Eq. (1).
6. The cluster points will enter into the cluster with which it has minimum Euclidean distance and it's consider as upper approximation of each cluster
7. Find the denominator. [Find euclidean distance between each object to all cluster center]. Which cluster center is minimum to each object that distance is consider as denominator].
8. Find the relative distance = [Find euclidean distance between each object to all cluster center]/denominator
9. Fix the cut off value = 1.5 and if greater than relative distance it indicates fuzziness. So cluster points become member of both the clusters.
10. The cluster points with fuzziness belong to upper approximation and cluster points with no fuzziness belong to lower approximation
11. Repeat the steps 3 to steps 10 until the cluster center becomes stable.

3.4.1 K-nearest neighbor

K-nearest neighbor algorithm works on non parametric systems for regression and classification. It identifies the membership in a particular class and also identifies the characteristics of a member also through regression.

- Fix with the nearest member value$K$
- Find the Euclidean distance between test case to all training case
- When K neighbors are similar a test case falls under the same category

The problem in this method is fixing up the K value because based on the value of K predicted value may get changed. Hence fix K values randomly until it gets saturated above which higher accuracy cannot be achieved. Here K = 3 is taken.

3.4.2 Decision tree

Works based on divide and conquer strategy. It was formulated by Quinlan in 1993. It works similar to trees. The procedure consists of the following steps.

– The information gain for all the variables are computed
– High information gain variable is selected as a root node
– Based on the root nodes characteristic a binary tree is constructed and construction stops when there is no variable
– Hence prediction labels are present in the leaf node.
– The rules are formed from the root to leaves and are similar to if-then rules.
– Testing cases are predicted based on the constructed decision rules

3.4.3 Random forest

Random forest trees work for classification, mainly regression and are constructed from many decision trees. Works similar to decision tree concepts. Thus the decision rules are formed from the above concept. Majority class decision is taken as output of the classification process.

3.4.4 Support vector machine

It is a supervised mining procedure used for regression, anomalies detection, and classification problems proposed by Boser, Vapnik, and Guyon in 1992. It works based on Gaussians radial kernel functions which can work on linear as well as non linear data sets. A support vector plane is computed which separates the members into two classes churners and non-churners. The decision lines can also be curve which can be achieved by support vector machine when it is a complicated task for many other classifiers. The two class members are divided into linearly separable members with an optimal decision line instead of highly complex curve like structure using mapping functions which are mathematical functions called kernels.

3.4.5 Naive Bayes

It is a simple probabilistic classifier works on Bayes Law. Mainly used in finding the posterior probability P(X/Y) obtained from prior probability [P(X)], Posterior probability condition works on Y [P(Y/X)] and [P(Y)] which is the prior probability of Y in Eq. (6).

$$P\,(X/Y) = \frac{P(Y/X)P(X)}{P(Y)} \tag{6}$$

4 Experiments and results

4.1 Experiment setup

The experiment consists of three phases. In phase I clustering is performed using rough k-means, k-means and fuzzy c-means algorithm then quality of the cluster is evaluated using Sum of Squared Error value (SSE). In Phase II the efficiency of the single classifier method like decision tree, k-nearest neighbor, random forest, support vector machine and naive bayes using precision, recall (sensitivity), specification, accuracy, and misclassification error. In Phase III hybrid performance of rough k-means algorithm and k-means algorithm with above stated classifiers are carried out.

Table 2 Confusion matrix for C3P

Actual	Predicted Churn	Predicted NonChurn
Churn	M_{11}	M_{12}
NonChurn	M_{21}	M_{22}

4.2 Performance measures

The performance measures that are used in this work are precision, recall (sensitivity), specification, accuracy, and misclassification error to analyze the comparative working of single classifiers and hybrid classifiers. The main aim of this work is to enhance accuracy and to reduce misclassification error. Table 2 shows the confusion matrix for credit card churn prediction model.

Where M_{11} represents the members who are already churn and found to be churned members, M_{12} misclassifies churned members as non-churned members, M_{21} misclassifies non-churn members as churn members, and M_{22} correctly classifies non-churn members as non-churn members. Precision, recall (sensitivity), specification, accuracy, and misclassification are calculated in Eqs. (7)–(11).

$$Precision = \frac{M_{11}}{M_{11} + M_{21}} \tag{7}$$

$$Sensitivity(Recall) = \frac{M_{11}}{M_{11} + M_{12}} \tag{8}$$

$$Specificity = \frac{M_{22}}{M_{21} + M_{22}} \tag{9}$$

$$Accuracy = \frac{M_{11} + M_{22}}{M_{11} + M_{12} + M_{21} + M_{22}} \tag{10}$$

$$Error = 1 - Accuracy \tag{11}$$

Sum of squared error (*SSE*) if sum of squared error is lesser value then the cluster members are well grouped into effective clusters. It indicates that the cluster members of a particular cluster are more similar to each other. Or if the value is greater the cluster members are not so similar even

Table 3 SSE comparison between rough k-means, k-means and fuzzy c-means

No of cluster's	Cluster algorithms		
	Rough K-means	K-means	FCM
K=2	4.8964e+05	5.4855e+05	1.5682e+05
K=3	4.4029e+05	5.4507e+05	1.4584e+06
K=4	4.4638e+05	1.0733e+06	1.3349e+06
K=5	5.4125e+05	1.3380e+06	1.8296e+06
K=6	4.4863e+05	5.6007e+05	1.5850e+06
K=7	5.0352e+05	1.3221e+06	1.4449e+06
K=8	4.0943e+05	1.3465e+06	1.0488e+06
K=9	6.7423e+05	1.0798e+06	1.8867e+06
K=10	4.2447e+05	1.8756e+06	1.5682e+06

they can be the members of different clusters highlighted in Eq. (11).

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} ||d_i - u_{ij}||^2 \tag{12}$$

where $||d_i - u_{ij}||^2$ = Euclidean distance computed between data points to cluster centers within a cluster.

4.3 Experiment 1

In experiment 1 list of unsupervised methods are analyzed comparatively using dominant algorithms like k-means, fuzzy c-means and rough k-means algorithm which works based on highly efficient rough set Theory. The system works based on clustering where customers with related view on the organization are clubbed or clustered together so that the churn or non-churn predictive classification has become possible. Here sum of squared error value for the three methods are comparatively and graphically studied. In this work X axis represents the number of clusters that are formed to group the customers and Y axis represents the SSE levels that are formed when we are changing the number of clusters and the same is analyzed for rough k-means system also. Table 3 and Fig 2 depict the SSE value for k-means clustering, fuzzy c-means clustering and rough k-means clustering algorithm.

From Figure 2, we can understand that rough k-means algorithm performs better with lower sum of squared values compared to k-means, fuzzy c-means algorithm. Total numbers of cluster members that are present inside each cluster are found that are produced by rough k-means, k-means and fuzzy c-means algorithm. Tables 4, 5, 6 depicts the total number of cluster members present inside each cluster when k=5. The Table 4 represent no of customer present inside each cluster and total number of all the cluster when using k-means algorithm. The Table 5 represent no of customer present inside each cluster and total number of all the cluster when using fuzzy c-means algorithm. Based on fuzzy membership function same customer is present in two or more clusters. The Table 6 represent no of customer present in lower approximation and upper approximation using rough k-means algorithm. Here lot of customer having fuzzification, that customer will present in both clusters.

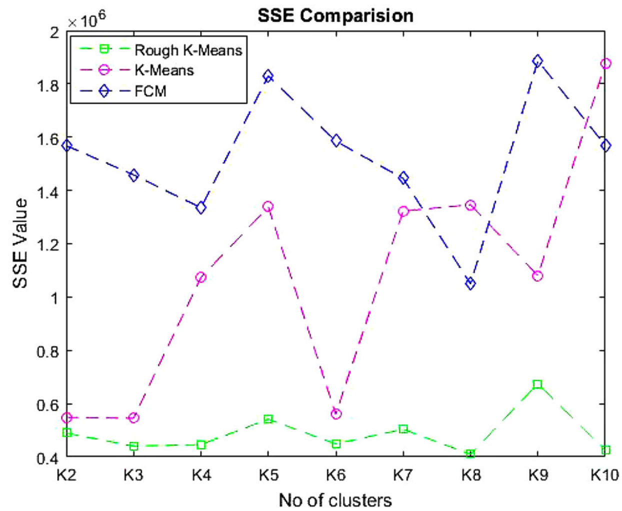

Fig. 2 SSE comparison between rough k-means, k-means and fuzzy c-means (Color figure online)

Table 4 Total number of cluster members present inside each cluster using k-means and cluster=5

Algorithm K-means	C1	C2	C3	C4	C5	Total
Customers	5310	6912	5100	6420	6258	30,000

Table 5 Total number of cluster members present inside each cluster using fuzzyc-means and cluster=5

Algorithm FCM	C1	C2	C3	C4	C5	Total
Customers	9812	5496	3488	621	10,633	30,050

Table 6 Total number of cluster members present inside each cluster using rough k-means and cluster = 5

Algorithm rough K-means	C1	C2	C3	C4	C5	Total
Customers in lower approximation	97	28	156	850	7	1138
Customers in upper approximation	582	936	14, 736	13, 746	0	30, 000
Customer in each cluster	679	964	14, 892	14, 596	7	31, 138

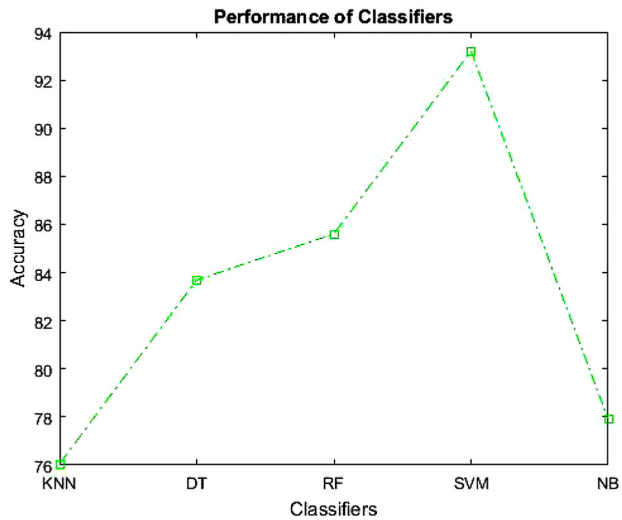

Fig. 3 Performance of five major classifiers using parameter accuracy

Table 7 The performance of classifier's

Classifier's	KNN	DT	RF	SVM	NB
Accuracy	76.05	83.70	85.60	**93.20**	77.90
Miss-error	23.95	16.30	14.37	06.76	22.10
Sensitivity	86.54	83.90	52.12	81.75	76.47
Specificity	07.08	09.54	24.33	05.67	11.22
Precision	85.96	84.20	81.90	66.27	80.60

Table 8 KNN along with K-means where cluster = 2– 6

K-means+KNN	C2	C3	C4	C5	C6
Accuracy	**94.08**	89.72	90.28	87.72	94.07
Miss-error	05.92	10.28	09.22	12.28	05.93
Sensitivity	99.05	97.57	97.92	96.85	99.05
Specificity	01.72	03.19	02.95	04.00	01.72
Precision	94.93	91.74	92.54	90.25	94.92

Table 9 DT along with K-means where cluster = 2– 6

K-means+DT	C2	C3	C4	C5	C6
Accuracy	90.68	86.19	86.28	**91.29**	90.79
Miss-error	09.32	13.81	13.72	08.71	09.21
Sensitivity	94.98	92.61	92.78	08.71	96.19
Specificity	08.79	15.73	14.51	05.15	05.71
Precision	95.20	92.35	92.29	94.94	94.14

4.4 Experiment 2

While using only one classifier the given data set can be partitioned into training data set and testing data set using hold-out process and efficiency of the system is measured using precision, recall (sensitivity), specificity, accuracy, and misclassification error. Of the five single classifiers models support vector machine found to perform better showing a maximum accuracy of 93.20%. Figure 3, Table 7 depicts the performance of five major classifiers considered in this work.

4.5 Experiment 3

4.5.1 Combining K-means clustering algorithm with five major classifiers

In this work we create a hybrid model combining k-means clustering algorithm with five major classifiers considered in this work. Initially our credit card churn prediction data set is clustered into relative clusters using k-means clustering algorithm and the clusters are divided into training data samples and testing data samples using hold-out technique. Based on training data sets the classifiers build a training model of the prediction system and uses that model to test the system. The major classifiers that are hybridized with k-means are decision tree, k-nearest neighbor, random forest, and support vector machine, naive bayes. The performance of the system is calculated in terms of five major performance measures like precision, recall (sensitivity), specificity, accuracy, and misclassification error. The support vector machine classifier performs better comparatively than other classifiers and it is found to increase the classification accuracy, reduces misclassification error to a large extent. Tables 8, 9, 10, 11, 12 depicts the performance of five major classifiers collaborated with k-means algorithm considered in this work. In Tables 7, 8, 9, 10, 11 which cluster group getting maximum accuracy value was indicated in bold.

4.5.2 Combining rough K-means clustering algorithm with five major classifiers

In the next hybrid model we combine single classifiers along with rough k-means clustering algorithm and the same exper-

Table 10 RF along with K-means where cluster = 2–6

K-means+RF	C2	C3	C4	C5	C6
Accuracy	90.78	92.11	87.31	**93.96**	86.58
Miss-error	09.22	07.89	12.69	05.04	13.42
Sensitivity	97.72	99.13	96.69	99.32	97.09
Specificity	01.98	00.00	03.04	03.17	00.00
Precision	92.73	92.86	89.96	95.57	88.89

Table 11 SVM along with K-means where cluster = 2– 6

K-means+SVM	C2	C3	C4	C5	C6
Accuracy	92.55	80.24	91.48	92.78	**95.26**
Miss-error	07.45	19.76	08.52	07.22	04.74
Sensitivity	99.09	92.02	98.33	98.58	99.21
Specificity	03.13	08.57	02.71	01.18	02.08
Precision	93.33	85.96	92.91	94.03	95.99

Table 12 NB along with K-means where cluster = 2– 6

K-means+NB	C2	C3	C4	C5	C6
Accuracy	89.96	86.31	86.52	84.88	**90.79**
Miss-error	10.04	13.69	13.48	15.12	09.21
Sensitivity	94.08	91.70	92.62	92.39	96.19
Specificity	10.06	13.62	16.60	15.95	05.71
Precision	95.30	93.47	92.71	90.98	94.14

Table 13 KNN along with rough K-means where cluster = 2– 6

Rough K-means+KNN	C2	C3	C4	C5	C6
Accuracy	**95.18**	72.40	90.50	90.06	87.40
Miss-error	04.82	22.60	09.50	09.94	12.60
Sensitivity	99.21	83.33	95.62	94.61	93.47
Specificity	00.00	13.33	09.62	04.00	13.06
Precision	95.91	83.85	94.36	94.88	92.74

Table 14 DT along with rough K-means where cluster = 2– 6

Rough K-means+DT	C2	C3	C4	C5	C6
Accuracy	91.60	**94.95**	86.19	84.35	90.79
Miss-error	08.40	05.05	13.81	15.65	09.86
Sensitivity	95.69	96.25	96.24	91.75	95.16
Specificity	03.45	00.00	06.00	13.62	05.50
Precision	95.93	98.66	89.10	91.04	94.44

Table 15 RF along with rough K-means where cluster = 2– 6

Rough K-means+RF	C2	C3	C4	C5	C6
Accuracy	90.86	**94.95**	93.63	87.46	90.61
Miss-error	09.14	05.05	66.36	18.54	09.39
Sensitivity	98.00	99.55	98.69	95.12	97.63
Specificity	02.49	03.03	02.63	06.09	01.82
Precision	92.55	95.36	94.81	84.82	92.63

Table 16 SVM along with rough K-means where cluster = 2– 6

Rough K-means+SVM	C2	C3	C4	C5	C6
Accuracy	94.62	88.25	91.44	91.80	**96.85**
Miss-error	05.38	11.75	08.56	08.20	03.15
Sensitivity	99.37	93.71	99.18	97.32	100.0
Specificity	02.04	08.88	0.00	00.00	01.54
Precision	95.18	93.73	92.13	94.17	91.72

Table 17 NB along with rough K-means where cluster = 2– 6

Rough K-means+NB	C2	C3	C4	C5	C6
Accuracy	91.18	88.11	**94.36**	81.34	93.46
Miss-error	08.82	11.89	05.64	18.36	06.54
Sensitivity	98.13	97.06	99.28	93.07	98.89
Specificity	02.28	04.05	02.35	08.11	00.00
Precision	02.28	90.48	95.01	86.69	94.44

imentation is carried out so that the performance is measured in terms of same measures like precision, recall (sensitivity), specificity, accuracy, and misclassification error. The experimentation results show that the entire single classifier algorithm performs better when combined with rough k-means algorithm than executing alone. Also hybrid model involving rough k-means clustering works better than k-means hybrid clustering algorithm. Also among all the hybrid models support vector machine combined with rough k-means shows greater accuracy, reduced misclassification error and lower specificity. Tables 13, 14, 15, 16, 17 depicts the performance of five major classifiers collaborated with rough k-means algorithm considered in this work. In Tables 12, 13, 14, 15, 16, 17 which cluster group getting maximum accuracy value was indicated in bold

4.5.3 *Performance comparison with single and hybrid classifier*

The new hybridized model works very well in predicting credit card churn compared to other model. It also works very well for other prediction systems and proved comparatively better results than the single classification methodologies that are tabulated in the following Table 18 and Fig. 4. Also the classifiers while working individually produced only less accuracy. But when combined with k-means clustering algorithm showed some improvement. In next hybrid model

Table 18 Accuracy comparison of single classifier and hybrid model's

Accuracy	KNN	DT	RF	SVM	NB
Single classifier	76.05	83.70	85.60	93.20	77.90
K-means hybrid model	94.08	91.29	93.96	95.26	90.79
Rough K-means hybrid model	95.18	94.95	94.95	96.85	94.36

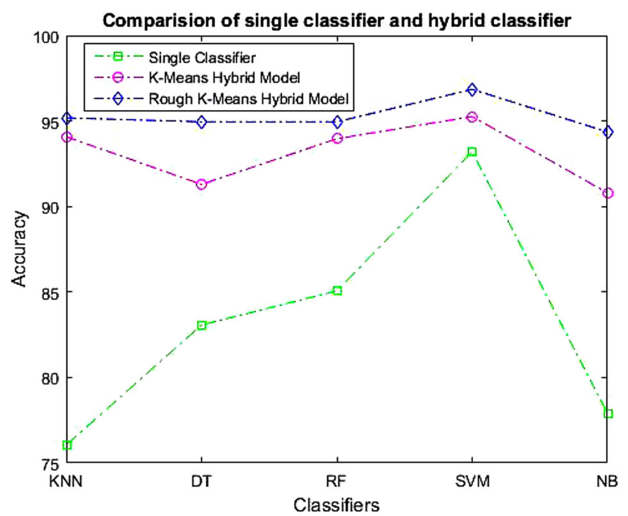

Fig. 4 Accuracy comparison of single classifier and hybrid model's (Color figure online)

classifiers when combined with rough k-means clustering algorithm provided higher accuracy. Among those hybrid models also support vector machine combined with rough k-means clustering algorithm works well with the highest accuracy and lesser error rate. In this work X axis represents the identified classifier's that are formed to classify the customer and Y axis represents the accuracy that are formed when we are changing the hybrid model and the same is analyzed for rough k-means hybrid system also. It also works very well for other prediction systems and proved comparatively better results than the existing methodologies that are tabulated in the following Table 19.

Table 19 Comparison of proposed method with existing approaches

Methods	Accuracy
Proposed rough K-means+ SVM	96.85
LR and DT by Nie [25]	73.45
RST+LS-SVM by Ning [5]	89.50
SVM by Zhao [6]	59.74
WK-means+FOIL by Ying [9]	93.88
Unsupervised+boosted DT by Bose [10]	91.52
KNN+LR by Yangming [16]	90.78

4.5.4 Testing proposed hybrid model for telecommunication data set

The proposed hybrid model rough K-means with SVM suitable for telecommunication data set also. The data set is obtained from Data Source Link http://www.sgi.com/tech/mlc/db/. [40]. The data set totally has 5000 member data and characterized as 20 various attributes and one churn prediction variable. Among 5000 members 707 members are identified as churners and 4293 members are identified as non-churners. Generally the dataset collected from any organization, or collected from websites or any other source if it is huge enough knowledge can be extracted to make decisions that will enhance the quality of the system in future. But the data collected will never be in good quality, it may have missing values, redundant data, noisy part, and wide range of differences between related data values. Hence it is the responsibility of miners to clean the data in the first step. Such cleaning process is called dataset preprocessing. Some unwanted features (State, Phone number) are deleted from dataset and reduced to 18, also string value features which can be converted to other numeric values which are not present in the data set. Finally apply the rough K-means +SVM hybrid model the performance is tabulated in Table 20.

Table 20 SVM along with rough K-means where cluster = 2–6 for telecommunication dataset

Rough K-means+SVM	C2	C3	C4	C5	C6
Accuracy	90.68	89.72	94.20	92.55	95.75
Miss-error	10.32	10.28	05.76	07.45	04.25
Sensitivity	94.98	97.57	79.67	99.09	98.79
Specificity	08.79	03.19	08.73	03.13	02.73
Precision	95.20	91.74	90.89	93.13	92.87

5 Conclusion

Finally as a conclusion this work focuses on building highly efficient hybrid data mining models. In this work UCI data set on credit card churn prediction is used to build hybrid models. Initially the data set has lot of variation in the data values so the samples are normalized. After preprocessing it the data set is divided into clusters using unsupervised methods like K-means and rough k-means algorithms. Thus the clusters that are obtained are divided into training and testing samples using hold-out process. Now the list of classifiers considered in this work carried out classification process. Of all the hybrid models support vector machine combined with rough k-means clustering algorithm works well with better accuracy. Thus the performance is measured in terms of precision, recall (sensitivity), specificity, accuracy, and misclassification error. The hybrid system works very well compared to single classifier model with high accuracy and lower misclassification error.

References

1. Yeh, I.C., et al.: The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Syst. Appl. **36**, 2473–2480 (2009)
2. Oyeniyi, A.O., Adeyemo, A.B., Oyeniyi, A.O., Adeyemo, A.B.: Customer churn analysis in banking sector using data mining techniques. Afr J. of Comp and ICTs. **8**(3), 165–174 (2015)
3. Farquad, M. A., et al.: Data mining using rules extracted from SVM: an application to churn prediction in bank credit cards. In: RSFDGrC '09 Proceedings of the 12th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. pp. 390–397 (2009)
4. Nie, G., et al.: Credit card churns forecasting by logistic regression and decision tree. Expert Syst. Appl. **38**, 15273–15285 (2011)
5. Wang, N., Niu, D. X.: Credit card customer churn prediction based on the RST and LS-SVM. IEEE conference on Service and systems and Service Management (2009)
6. Zhao, J., et al.: Bank customer churn prediction based on support vector machine: taking a commercial bank's VIP customer churn. IEEE Conference on Wireless Communication, Networking and Mobile computing (2008)
7. Zhao, Y., et al.: Customer churn prediction using improved one-class support vector machine. DMA, LNAI 3584, pp. 300–306. Springer-Verlag Berlin Heidelberg, Berlin (2005)
8. Shao, H., et al.: Construction of Bayesian classifiers with GA for predicting customer retention. IEEE Conference on Natural Computation (2005)
9. Huang, Ying, Huang, Y., Kechadi, T.: An effective hybrid learning system for telecommunication churns Prediction. J. Expert Syst. Appl. **40**, 5635–5647 (2013)
10. Bose, I., Chen, X.: Hybrid models using unsupervised clustering for prediction of customer churn. J. Organ. Comput. and Electron. Commer. **19**, 133–151 (2009)
11. Huang, B.Q., Kechadi, M.T., Buckley, B.: Customer churns prediction in telecommunications. J. Expert Syst. Appl. **39**(1), 1414–1425 (2012)
12. Khashei, M., Hamadani, A.Z., Bijari, M.: A novel hybrid classification model of artificial neural networks and multiple linear regression models. J. Expert Syst. Appl. **39**(3), 2606–2620 (2012)
13. Tsai, C.F., Lu, Y.H.: Customer churns prediction by hybrid neural networks. J. Expert Syst. Appl. **36**(10), 12547–12553 (2009)
14. Yeshwanth, V, Raj, V.V., Saravanan, M.: Evolutionary churn prediction in mobile networks using hybrid learning. In: Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference, (FLAIRS), Palm Beach, Florida, USA, May 18–20. AAAI Press (2011)
15. Hung, S.Y., et al.: Applying data mining to telecom churn management. J. Expert Syst. Appl. **31**(3), 515–524 (2006)
16. Zhang, Y., et al .: A hybrid KNN-LR classifier and its application in customer churn prediction IEEE conference on system, man and cybernetics (2007)
17. Lee et al,.: Customer churn prediction by hybrid model. Advance data mining and application Springer (2006)
18. Kawale, J., Pal, A., Srivastava, J.: Churn prediction in MMORPGs: a social influence based approach. Proc. 2009 Int. Conf. Comput. Sci. Eng. **4**, 423–428 (2009)
19. Suznjevic, M., Stupar, I., Matijasevic, M.: MMORPG player behavior model based on player action categories. Proceedings of the 10th Annual Workshop on Network and Systems Support for Games. IEEE Press (2011)
20. Chen, K. T., Lei, C. L., Network game design: hints and implications of player interaction. In: Proceedings of 5th ACM SIGCOMM Workshop on Network and System Support for Games. pp. 1–9 (2006)
21. Soeini, R.A., Rodpysh, K.V.: Applying data mining to insurance customer churn management. Int. Proc. Comput. Sci. Inf. Technol. **30**, 82–92 (2012)
22. Verbeke, W., Martens, D., Baesens, B.: Social network analysis for customer churn prediction. Appl. Soft Comput. **14**, 431–446 (2014)
23. Malik, Z.K., Hussain, A., Jonathan, W.: An online generalized eigenvalue version of Laplacian Eigenmaps for visual big data. Neurocomputing **173**, 127–136 (2016)
24. Anil Kumar, D., Ravi, V.: Predicting credit card customer churn in banks using data mining. Int. J. Data Anal. Tech. Strateg. **1**(1), 4–28 (2008)
25. Nie, G., et al.: Finding the hidden pattern of credit card holders churn: a case of China, ICCS 2009 part II. LNCS **5545**(1), 561–569 (2009)
26. Wang, G., et al.: Predicting credit card holder churn in banks of China using data mining and MCDM. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (2010)
27. Chiang, D.A., et al.: Goal-oriented sequential pattern for network banking churn analysis. Expert Syst. Appl. **25**, 293–302 (2003)
28. Jinbo, S., et al.: The application of AdaBoost in customer churn prediction. IEEE Conference on Service and Systems and Service Management (2007)
29. Mutanen, T., et al.: Customer churn prediction: a case study in banking. ECML/PKDD 2006 Workshop on Practical Data Mining: Applications, Experiences and Challenges (2010)
30. Olle, G., Hybrid, A.: Churn prediction model in mobile telecommunication industry. Int. J. e-Educ. e-Bus. e-Manag. Learn. **4**(1), 55–62 (2014)
31. Peters, G.: Some refinements of rough k-means clustering. Pattern Recognit. **39**, 1481–1491 (2006)
32. Kim, K., Jun, C.-H., Lee, J.: Improved churn prediction in telecommunication industry by analyzing a large network. Expert Syst. Appl. **41**(15), 6575–6584 (2014)
33. Dataset. https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients
34. Lingras, P., West, C.: Interval set clustering of web users with rough k-means. J. Intell. Inform. Syst. **23**, 5–16 (2004)
35. Kumar, P.: Comparative study of K-means, pam and rough K-means algorithms using cancer datasets. 2009 International Symposium on Computing, Communication, and Control (ISCCC 2009)
36. Lingras, P., West, C.: Interval set clustering of web users with rough kmeans, Technical Report 2002–002. St. Marys University, Halifax, Canada, Department of Mathematics and Computer Science (2002)
37. Parmar, D., et al.: MMR: an algorithm for clustering categorical data using rough set theory. Data Knowl. Eng. **63**(2007), 879–893 (2007)
38. Amin, A., et al.: Churn prediction in telecommunication industry using rough set approach. New Trends in Computational Collective Intelligence, pp. 83–95. Springer International Publishing, Berlin (2015)

39. Amin, A., et al.: Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing **237**, 242–254 (2017)
40. Data Source Link. http://www.sgi.com/tech/mlc/db/

Dr. R. Rajamohamed received B.E in Electronics and Communication Engineering at Madurai Kamaraj University, Madurai, Tamilnadu, India, in the year 1999, M.E in Applied Electronics at Madurai Kamaraj University, Madurai, Tamilnadu, India, in the year 2001 and Ph.D. degree in Electronics and communication Engineering (Wireless Communication) from University of St. Peter University, India, in 2016. He started his career as lecturer in RVS College of Engineering and Technology, Dindigul, Tamilnadu, India in the year 2001. Now he is working as the Associate Professor, in the department of Electronics and Communication Engg., Indra Ganesan College of Engg., Tiruchirappalli. During the year 2011-2017, he has worked as Electronics and Instrumentation engineer in Saudi Arabia. His area of interest includes Mobile ad hoc networks, simulation and network security and network on chips, cluster computing, optical communication.

J. Manokaran received B.E in Electronics and Communication Engineering at Noorul Islam College of Engineering, Kanyakumari , Tamilnadu, India, in the year 2009, M.E in Applied Electronics at Anna University of Technology, Tirunelveli, Tamilnadu, India, in the year 2011. Now he is working as the Assistant Professor, in the department of Electronics and Communication Engg., Indra Ganesan College of Engg., Tiruchirappalli. His area of interest includes Data mining, Image processing, Signal processing, cluster computing, Big data analytics