



# A NOVEL APPROACH TO SUSTAIN CUSTOMER LOYALTY IN BANKING SECTOR USING MINING TECHNIQUES

<sup>1</sup>R. Boopathi, <sup>2</sup>Dr. B. Umadevi

<sup>1</sup>Research Scholar, <sup>2</sup>Assistant Professor & Head

<sup>1</sup>P.G. & Research Department of Computer Science

<sup>1</sup>Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India

<sup>2</sup>P.G. & Research Department of Computer Science

<sup>2</sup>Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India

---

## Abstract

Today all commercial sectors are running in hectic competition. At the same time giving the best services and customer retention is a great challenge also. In considering the various influencing factors the banking groups need to put a potential mechanism in customer retention. The concept of customer loyalty has received much consideration and attention in different industries. It is more important to build consumer loyalty is seen as the key factor in winning market share and developing a sustainable competitive advantage. Banking industry has no exception as it has high interaction with the customers, since managers need to understand influencing factors of the customers towards their respective banks. This paper tries to find the factors of customer loyalty and their relationships with the banking industry. The relationships of different factors with each other are also studied to keep as well as to maintain the quality service to the customers.

**Keywords:** Loyalty, Retention, CRM, K-means, DBSCAN.

---

## 1. Introduction

The revolution in technological factors has made tremendous change in financial institutions offers remarkable services in banking sectors. The business strategy of every financial service company is the ability to retain existing customer and reach new prospective customers. The banking industry is growing rapidly to keep pace with the growth of the industry through scientific advancements and pioneering ideas to promote the organization. A group of products and services are offered by bank in a diversified manner in order to maintain the customer retention. Accumulation of operational data inevitably follows from this growth in industry. There exists an increasing need to convert their data into a corporate asset in order to stay ahead and gain a competitive advantage.

Data mining plays vital role in accordance with bank related activities. In the financial area, data mining has been [1] applied successfully in determining the likely eligible candidate for loan disbursement, finding profitable customers, products, characterizing different product segments. All of these factors are taxing replaces traditional polices and business approaches forcing banks to consider reinventing themselves to win in the marketplace. Nowadays it is difficult to find customers for loan disbursement are really a challenging issue in the banking era [2]. This paper primarily focuses to maintain the customer loyalty by the banking sectors.



## 1.1 Customer Relationship Management

The Customer Relationship Management (CRM) was initiated since 1990. Expansion of relationship with customers may lead to loyalty and retention. Customer relationship management has various faces such as 1) Customer identification 2) Customer attraction 3) Customer retention 4) Customer development. These dimensions are called the closed-loop customer relationship management. CRM is the blend of processes from customer draw to maintain the data mining techniques to exploit customer value. CRM ensures that a minor change in retention rate would lead to significant market share [3]. The first two dimensions of CRM, customer identification and attraction are costly and confirm the monetary benefits of companies.

Customer loyalty becomes very important for an organization to retain its current customers. Gaining customer loyalty becomes a key objective for majority banking organizations in the world. In overall, a successful strategic can help the organization to retain the customer in order to survive in highly competitive market. Developing an efficient customer loyalty strategic is very important to organizations especially banks to build a strong relationship with their customers.

In this paper, a new approach has been proposed to maintain the customer loyalty by the banking sector. It may lead to make sanctioning loans based on different parameters. The paper is structured as follows. The section one describes about introduction. Section 2 deals about background study and its related works. The methodology of the research work is explained in section three. In section four portrays the Experiment results. Finally the paper is concluded in last section.

## 2. Background and Related Works

Data mining has emerged [4] over recent years as an extremely powerful approach to extracting meaningful information from large databases and data warehouses. Since the increased computerization of business transactions, improvements in storage and processing capacities of computers, as well as significant advances in knowledge discovery algorithms, those all have contributed to the evolution of the data mining.

### 2.1 Literature Review

In this paper, intentional approaches using various data mining techniques are collected to analyse the customer's loyalty at various levels. The study related to data mining for extracting and predicting the customer's loyalty used in various models and the comprehensive literature review of various researcher's works are stated below:

A.O.Oyeniya and A.B.Adeyemo *et al.*, presented a data mining model that can be used to predict which customers are most likely to churn (or switch banks) [5]. Their study used real-life customer records provided by a major Nigerian bank. The raw data was cleaned, pre-processed and then analysed using WEKA, a data mining software tool for knowledge analysis. Simple K-Means was used for the clustering phase while a rule-based algorithm, JRip was used for the rule generation phase. Their results showed that the methods used can determine patterns in customer behaviours and help banks to identify likely churners and hence develop customer retention modalities.

Yogita Bhaskar and Ajit *et al.*, focused on study of loan applicants by using clustering [6] approach. They used and analysed the performance of K-means algorithm, hierarchical clustering algorithm and EM *i.e.*



expectation maximization algorithm was measured. The WEKA tool was used to predict the performances of algorithms are compared on the basis of accuracy and running time.

Dr. V.P.Eswaramurthy and S.Induja *et al.*, presented light on the underlying [7] technologies and the viewpoint applications of data mining in predicting the churn behaviour of the customers and hence paving way for better Customer Relationship Management.

Dr. Mohammad Miyan *et al.*, presented the applications of data mining [8] in the banking sectors. It contains a general overview of data mining, providing a definition of the concept, primary data mining techniques and mentioning the main fields for which the data mining can be applied. He also presented the banking business sector which can benefit from the use of DM tools, along with their use cases *i.e.*, retail and insurance banking sector.

Mahin Tasnimi *et al.*, focused on data mining models in the field of clustering, [9] to categorize customers to improve customer relationship management and marketing strategies for each category of customers. The RFM variables are used to category Customers and the data collected using the software SPSS Clementine was analyzed and K-means algorithm is used to cluster clients. Finally, using decision tree algorithm rules for each category of customers are extracted and the accuracy of the model was evaluated by the software.

K.Vanitha and G.Roch Libia Rani *et al.*, investigated the performance of different classification and clustering algorithms [10] using weka software. The J48, Naive Bayes and Simple CART Classification algorithms are evaluated based on accuracy, time efficiency and error rates. The K-means, DBScan and EM clustering algorithms are evaluated based on accuracy of clustering.

### 3. Methodology

Today Data mining becomes the most promising research area. Data mining [11] supports set of techniques that can be used to extract relevant and interesting knowledge from data [12]. Data mining techniques such as association, rule mining, classification, clustering and prediction.

Clustering is a data mining technique used for making group of abstract objects into classes of similar objects are grouped into one cluster and dissimilar objects are grouped in another

Cluster. Clustering is a data mining technique have its considerable amount of research in science, information technology field, medical science, image processing, document classification, clustering analysis is also used in banking industries also. Our research draws the power of clustering techniques such as K-means algorithm and DBSCAN (density based) algorithms for partitioning the data. The Classification algorithm J48 (C4.5) is applied over the clustered data to predict the loyalty of the customer in banking sectors. The step involved in predicted [13] the loyalty behaviour of customer which is given in Figure 1. The algorithms are discussed below:

#### 3.1 K-means Clustering

The K-means algorithm is one of the clustering algorithms whose use is widespread. K-means, which belongs to the class of partitioning algorithms, has two main advantages: it is very easy to implement and it takes little time to run, which makes it suitable for large data sets. The K-means is the most common partitioned clustering algorithm which is used to partition  $n$  observations into  $K$  clusters in which each observation belong to cluster with nearest mean. It is simple, non supervised iterative learning method. The idea behind classifying set of data objects into  $K$  number of clusters where  $K$  is fixed initially. It first fixes initial group centroids [14].

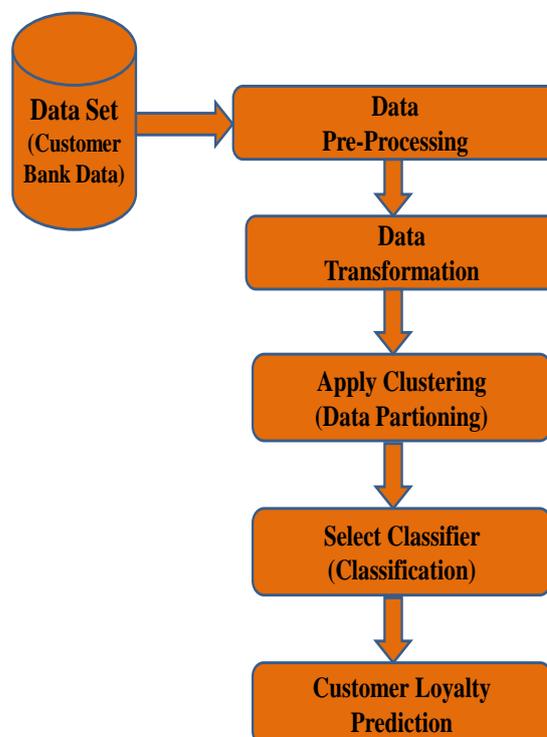
Then assign each object to the group that has closest centroid. Once all the objects are assigned it recalculate positions of centroids. Again repeat the same process until centroids not change.

### 3.2 DBSCAN

DBSCAN (Density Based Spatial clustering of application with noise) is to create clusters with minimum size and density. Density is defined as the minimum number of points within a certain distance of each other. It requires two parameters: epsilon (eps) and minimum points (minPts). DBSCAN does not require you to know the number of clusters in the data a priori. [15]. DBSCAN does not have a bias towards a particular cluster shape or size. DBSCAN is resistant to noise and provides a means of filtering for noise [16] if desired. DBSCAN does not respond well to high dimensional data. As dimensionality increases, so does the relative distance between points making it harder to perform density analysis. DBSCAN does not respond well to data sets with varying densities.

### 3.3 J48 (C 4.5)

The J48 (C4.5) technique is one of the decision tree families that can produce both decision tree and rule-sets and construct a tree for the purpose of improving prediction accuracy. The J48 classifier is among the most popular and powerful decision tree classifiers. The J48 creates an initial tree using the divide-and-conquer algorithm.



**Figure 1:** Steps involved in Predicting Loyalty of Customers in Banking Sectors



#### 4. Experiment Results

The data sets of different customers are collected from various banking sectors located at sivagangai. At the outset a sample of 200 took for analysis with different attributes mentioned below. The attributes are Customer-ID, Gender, Age, Transaction-Duration, Types-Of-Loans-Availed, Income, Loan-Eligibility and Loyalty. The acronym for the attribute is listed in Table 1.

The data set is pre-processed; The K-Means and DBSCAN algorithms are applied over the data to partitioning the data into ( $K = 5$ ) different clusters. The Random initialization method and the ten as the initial seed value are used to form the clusters. In K-Means, the Euclidean Distance and in DBSCAN, the Manhattan Distance metrics are used to measure distance between of an observation and the initial cluster centroids. From the cluster data, the customer behaviour is analysed through the classification technique J48 (C4.5) to extract the loyalty of customer's in banking sectors. In our analysis the "Loyalty" is used as the class label to predict or classify the different types of customers. The outcomes of the results are used to identify or categorize the different types of customers. The interpretations are given in Figures 2– 5.

Table 1: Attribute Description

Attribute-Name	Type	Possible-Values
Customer-ID	Numeric	Range : 1-200
Gender	Nominal	female, male
Age	Numeric	Range : 18-58
Transaction-Duration	Numeric	Range : 2 - 30
Types-Of-Loans-Availed	Nominal	PL - Personal-Loan HL - Housing-Loan CL - Car-Loan ED - Education-Loan JL - Jewel-Loan
Income	Numeric	Range 20,000-1,00000
Loan-Eligibility	Nominal	PL - Personal-Loan HL - Housing-Loan CL - Car-Loan ED - Education-Loan JL - Jewel-Loan
Loyalty	Nominal	low,high,medium, very low, very high

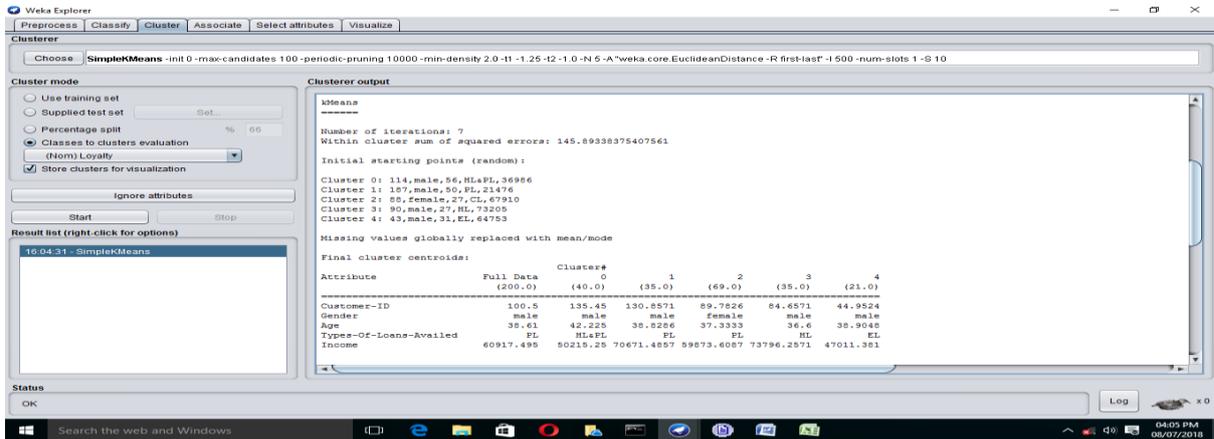


Figure 2 : Clustering by K-means

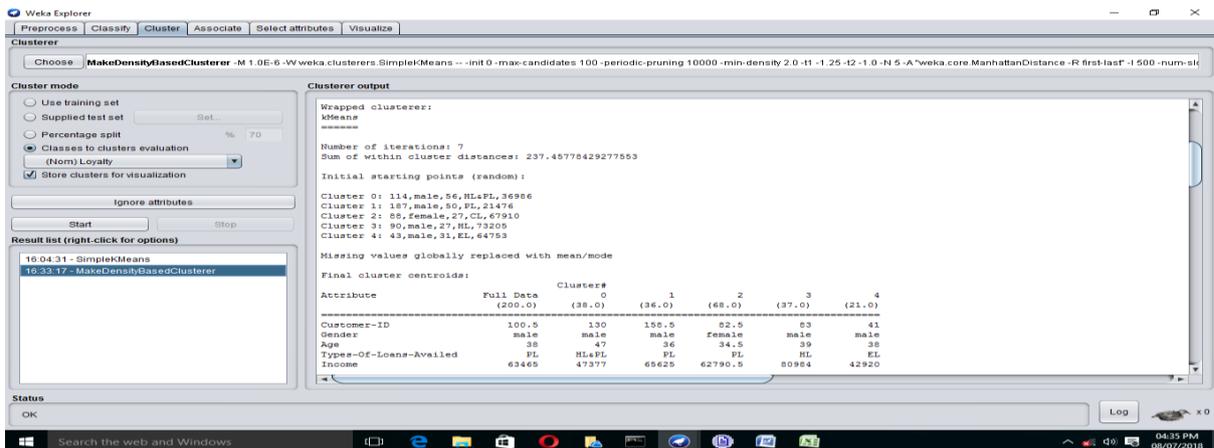


Figure 3 : Clustering by DBSCAN(K-means with density based)

Table 2 : Clustering by K-Means

Cluster -Name	Customer -ID	Gender	Age	Types-Of-Loans -Availed	Income	Loyalty
Cluster-0	135.45	male	42.225	HL&PL	50215.25	medium
Cluster-1	130.8571	male	38.8286	PL	70671.4857	verylow
Cluster-2	89.7826	female	37.3333	PL	59873.6087	high
Cluster-3	84.6571	male	36.6	HL	73796.2571	veryhigh
Cluster-4	44.9524	male	38.9048	EL	47011.381	low

Table 3 : Clustering by DBSCAN(K-Means with density based)

Cluster-Name	Customer-ID	Gender	Age	Types-Of-Loans-Availed	Income	Loyalty
Cluster-0	130	male	47	HL&PL	47377	medium
Cluster-1	158.5	male	36	PL	65625	verylow
Cluster-2	82.5	female	34.5	PL	62790.5	high
Cluster-3	83	male	39	HL	80984	low
Cluster-4	41	male	38	EL	42920	veryhigh

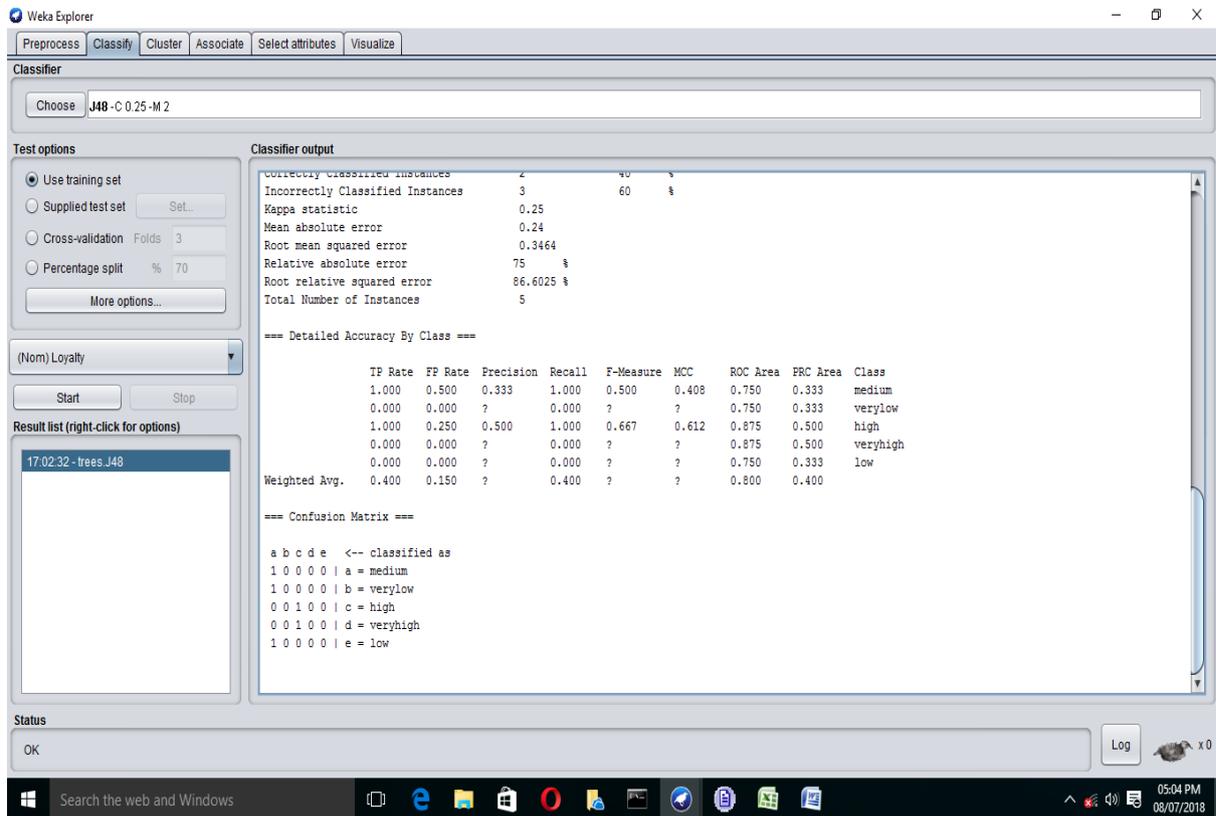


Figure 4 : Predicting Customer's Loyalty by J48 (C4.5)

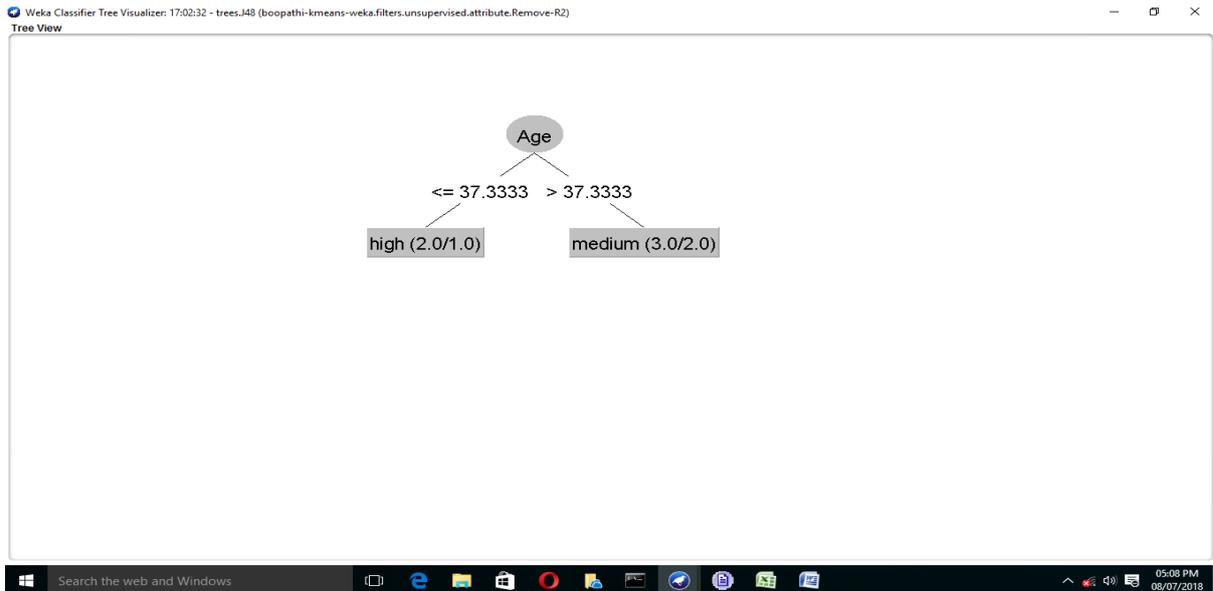


Figure 5 : Predicting Customer by Loyalty using J48 (C4.5)

## 5. Conclusion

The data has been collected from various banking sectors located in sivagangai. The research objective is in categorizing the customers and their relationship with banks. The analysis primarily addresses that how the banking sectors constantly keeps on relationship with their customers. It is determined by their regularity or non defaulters. The behavioural attributes are applied through the clustering algorithms to segment the customers. From the clusters the J48 algorithm is applied to predict the customer loyalty and customer retention. It is highly justifies that customers with age less than or equal to 37 holds high loyalty and others who are greater than 37 are possessing medium loyalty. From that the Banking sectors are highly motivated to sanction loans to their customers with such loyalty parameters.

## References

- [1] Wesley Chu and T.Y. Lin , January 2005 ,“A Data Mining Approach for Retailing Bank Customer Attrition Analysis: Special Issue: Foundations and Advances in Data Mining”, *Article in Applied Intelligence* ,22(1) , DOI: 10.1023/B:APIN.0000047383.53680.b6 ,Source: dx.doi.org.
- [2] Boopathi.R and Umadevi .B.Dr, April 2018 ,” Retention of Customer Loyalty in Banking Sector Through Mining Techniques: A Review”, *International Journal of Creative Research Thoughts (IJCRT)*, Volume 6, Issue 2 ISSN: 2320-2882.
- [3] Mohammad Ali Afshar Kazemi, Negar Estemdad and Alireza Poorebrahimi, March 2015,” Customer Behavior in Banking Industry: Comparison of Data Mining Techniques”,*Journal of Advanced Management Science* Volume. 3, No. 1, ©2015 Engineering and Technology Publishingdoi: 10.12720/joams.3.1.13-16.



R. Boopathi *et al*, International Journal of Computer Science and Mobile Applications,  
Vol.6 Issue. 7, July- 2018, pg. 13-22

ISSN: 2321-8363

UGC Approved Journal

Impact Factor: 5.515

- [4] Chih-Fong Tsai and Yu-Hsin Lu, 2010, "Data Mining Techniques in Customer Churn Prediction", *Recent Patents on Computer Science*, Volume 3, 28-32.
- [5] Oyeniyi A. O. and Adeyemo A.B., September 2015, "Customer Churn Analysis In Banking Sector Using Data Mining Techniques", *African Journal of Computing & ICT*, Vol 8. No. 3. Pp 165-174.
- [6] Yogita Bhapkar and Ajit More Dr., 2016, "Comparative analysis of clustering algorithms for the study of home loan applicants using WEKA tool", *Sinhgad Institute of Management & Computer Application (SIMCA)*, ISBN:978-93-5254-640-4.
- [7] Eswaramurthy Dr.V.P and S.Induja, 2014," A Study on Customer Rentention using Predictive Data mining Techniques", *International Journal of Computer & Organization Trends (IJCOT)*, Volume - 4 Issue – 5.
- [8] Miya and Mohammad, Jan/Feb2017, "Applications of Data Mining in Banking Sector", *International Journal of Advanced Research in Computer Science*, Vol. 8 Issue 1, p108-114. 7p.
- [9] Mahin Tasnimi, 2016, "Categories Customers Using Data Mining for Effective Communication and Increased Profitability", *International Academic Journal of Science and Engineering*, Volume. 3, No. 6pp. 199-209.
- [10] Roch Libia Rani .G and Vanitha. K, Nov. –Dec, 2010, "Analysis Of Classification And Clustering Algorithms Using Weka For Banking Data", *International Journal of Advanced Research in Computer Science*, Volume. 1 (4), 104-107.
- [11] Umadevi B. and Sundar D, Alli Dr P., January 2013, "An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50", *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, Vol.3.
- [12] Umadevi B, Sundar D., Alli Dr.P., 10th June 2014, "Novel Framework For The Portfolio Determination Using PSO Adopted Clustering Technique", *Journal of Theoretical and Applied Information Technology*, Vol. 64- No.1.
- [13] Umadevi B.and Sundar D., Alli Dr.P., 2013, "An Optimized Approach to Predict the Stock Market Behavior and Investment Decision Making using Benchmark Algorithms for Naive Investors", *Computational Intelligence and Computing Research (ICCIC)*, *IEEE International Conference*, 26-28 Dec. 2013, Page(s):1 – 5978-1-4799-1594-IINSPEC Accession Number: 14061140.
- [14] Shailendra Singh Raghuvanshi and PremNarayan Arya, November – December 2012, "Comparison of K-means and Modified K-mean algorithms for Large Data-set", *International Journal of Computing, Communications and Networking*, Volume 1, No.3, ISSN 2319-2720.
- [15] Maria Halkidi, Yannis Batistakis and Michalis Vazirgiannis, 2001, "On Clustering Validation Techniques", *Journal of Intelligent Information Systems*, Volume, 17:2/3, 107–145.
- [16] Boriana L. Milenova and Marcos M. Campos, 2002," O-Cluster: Scalable Clustering of Large High Dimensional Data Sets", *Oracle Data Mining Technologies*.



R. Boopathi *et al*, International Journal of Computer Science and Mobile Applications,  
Vol.6 Issue. 7, July- 2018, pg. 13-22

ISSN: 2321-8363

UGC Approved Journal

Impact Factor: 5.515

## **Authors' Biography**

**R. BOOPATHI** is an M.Phil Research Scholar in PG & Research Department Of Computer Science, Raja Doraisingam Government Arts College, Sivaganga, Tamilnadu, India. Her research concentration include in Data mining, Machine Learning and its applications. She also published papers in the international journal.



**Dr. B. UMADEVI** has received her Doctoral degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, India. Currently working as Assistant Professor & Head- P.G and Research Department of Computer Science, Raja Doraisingam Government Arts College, Sivagangai-Tamilnadu, India. She has over 23 years of Teaching Experience and published her research papers in various International, National Journals and Conferences. Her research interests include Data Mining, Soft Computing and Evolutionary Computing. She got the Best Paper Award for her publication in the IEEE International Conference on Computational Intelligence and Computing Research held on 27<sup>th</sup> Dec 2013 at VICKRAM College of Engineering and Technology.

