C. Sujitha *et al*, International Journal of Computer Science and Mobile Applications,
Vol.6 Issue. 7, July- 2018, pg. 23-33 **ISSN: 2321-8363**
UGC Approved Journal
Impact Factor: 5.515

AN ANALYTICAL APPROACH FOR CUSTOMER RETENTION THROUGH BUSINESS INTELLIGENCE

[1]C. Sujitha, [2]Dr. B. Umadevi
[1]Research Scholar, [2]Assistant Professor & Head
[1]P.G. & Research Department of Computer Science
[1]Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India
[2]P.G. & Research Department of Computer Science
[2]Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India

Abstract
Today the Business community needs a sophisticated environment for analyzing in order to make an opt decision. Business Intelligence (BI) applies various strategies and techniques that will help enterprises to perform data analysis effectively. BI also provides the different views such as legacy, current and futuristic to incorporate the business strategies. In general every business corporate will attract their customer either seasonally or periodically. The primary factor for products promotions the business sector needs data in relevancy with choices, likings and disliking of their valuable customers. There are various factors that influence the product promotion, even though the considerations are being given to only limited parameters. This research paper focuses the ideological analysis to identify the likings and disliking supporting for the business community for appropriate decision making.

Keywords: Business Intelligence (BI), likings, disliking and promotion, MLP.

1. Introduction

The promotional activities of any business are based on psychological factors of every human being. It is very interesting to study of the person who wants to buy certain product from shopping place, at the time of purchasing the product, why does he want to buy it?. This makes interest on the study about customers. The analytical data will be converted into statistical format which supports for interpret the buying behavior of each customer. The study shows that how the organizations to have to consider their marketing strategies by understanding the following issues:

(i) The psychology of each customer such as: feel over the product, reasons for selecting different alternative products and reason for rejecting other items.
(ii) The environment factor that influence each customer on multi products.
(iii) The strategy behind each consumer while shopping or making other marketing decisions.
(iv) What is the logic behind motivation, importance or interest between products entail differs in customer selection and
(v) It is a great challenge for the management to improve their marketing campaigns more effectively to reach the customer.

 ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

Data mining techniques are applied on current data and historic data which may be in large volumes to find out the hidden knowledge. Upon collecting the data from various sources, it is preprocessed, validated and stored in Data Warehouse/Data Marts. BI tools extract required information from this data and use for knowledge discovery process. Data Mining can be done using classification, clustering Market Basket analysis, Association rule mining, and link based analysis and so on.

In this paper, the customer behavior analysis and the key machine learning algorithms that have been used to solve it are discussed. 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 portraits the Experiment results. Finally the paper is concluded in last section.

2. Background and Related Works

Data mining (DM) can used to understand the consumer buying behavior using various techniques more than a decade. Every activity of a consumer in a supermarket is treated as a byte of data in data mining world. The consumer attributes related to supermarket such as how he/she spends, which day, what time normally he/she does the shopping, what they buy most often, how much they buy, in that locality etc.

Data mining is the process of finding correlations [1] or patterns among dozens of fields in large relational databases. Data mining is primarily used day by day comprise with a strong consumer focus retail, financial, communication & marketing organizations. It enables these companies to determine relationships among internal factors such as price, product positioning [2] or staff skills & external factors such as economic indicators, competition & customer demographics.

Classification techniques are supervised learning techniques [3] that classify data item into predefined class label. It is one of the most useful techniques [4] in data mining to build classification models from an input data set. The classification techniques commonly build models that are used to predict future data trends.

2.1 Literature Review

In this paper, the diversified approach using a variety of data mining techniques are collected to analyse the consumer behavior analysis in various business levels. The study related to data mining for extracting and predicting the consumer behavior used in a different data models and the comprehensive literature review of various researcher's works are discussed below.

Nan-Chen Hsieh and Kuo-Chung Chu presented a two-stage [5] frame-work of consumer behavior analysis, and the key feature is a cascade involving self-organizing map (SOM) neural network to divide customers into homogeneous groups of customers and a decision-tree simplified method to identify relevant knowledge. Identifying consumers by their approach is helpful characteristic of customers and facilitates marketing strategy development.

Liu Hongyan and Liu Zhenyu progressed of [6] theoretical research services and service quality management theory, according to a large data analysis and complex network, is recommended by consumer behaviour.

Monica LIA presented a customer data analysis model in a telecommunication company [7] and business intelligence tools for data modelling, transforming, data visualization and dynamic reports building. They also explored, a mature market, knowing the information inside [8] the data and making forecast for strategic

decision become more important in Romanian Market. Business Intelligence tools are used in business organization as support for decision making.

Abhijit Raorane and R.V.Kulkarni elucidated to know consumer behaviour, his psychological condition at the time of purchase and how suitable [9] data mining method apply to improve conventional method. Moreover, in their experiment, association rule is employed to mine [10] rules for trusted customers using sales data in a super market industry.

T.V. Rajinikanth and D Vasumathi intended at using both BI and DM techniques together in visualizing and analyzing the business data. Explorative data visualizations were also made to identify the hidden patterns for effective decision making suitable to their needs.

Filipe Pinto, Pedro Gago and Manuel Filipe Santos developed a systematic approach for the use of DM techniques as a new paradigm in Business Intelligence in DBM projects, considering analytical and marketing aspects. A cross-table was proposed to associate DBM activities to the appropriate DM techniques. Their framework guided the development of DBM projects, contributing to improve their efficacy and efficiency.

Gaurang Panchal, Amit Ganatra, Y P Kosta and Devyani Panchal discussed behavioral analysis of different number of hidden layers and different number of hidden neurons [11]. What to do while neural network is not getting train or errors are not getting reduced. They used Neural Network to reduce errors perform correct classification.

3. Methodology

This section presents detailed description of various data mining [12] algorithms that have been adopted to identify customer behaviour. Brief overviews of Decision tree methods such as J48 and Random Tree are presented in sections 3.1 and 3.2 respectively. The Multilayer Perception method is also discussed in section 3.3.

3.1 J48 (C 4.5)

Decision trees are one of the predictive [13] modeling approaches extensively used in [14] data mining, where in a tree is used to explicitly represent decisions and decision making [15]. These are the structured regression models. The goal of decision tree is to create a model that predicts [16] the value of a target based on several input variables. Each node of a decision tree represents one of the traffic usage attributes of the customer. Leaves represent class labels and branches represent conjunctions of features that lead to class labels [17].

The decision tree used in J48 is specifically chosen for its pruning ability and exceptional handling of missing classes, which no other tree could perform. 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.

3.2 Random Tree

Random Tree algorithm is a popular tree based ensemble learning technique. The bagging type used here is "ensembling". In bagging, successive trees do not depend on earlier trees each is independently [18] constructed using a different bootstrap sample of the data set. In the end, a simple majority vote is taken for prediction. Random Trees are different from standard trees in that for the latter each node is split using the best

split among all variables. In a random tree [19], each node is split using the best among a subset of predictors randomly chosen at that node. This additional layer of randomness makes it robust against over-fitting.

3.3 Multilayer Perceptron

Multi Layer Perceptrons (MLPs) are feed forward neural networks trained with the standard back propagation algorithm. They are supervised networks [20] so they require a desired response to be trained. They learn how to transform input data into a desired [21] response, so they are widely used for pattern classification. With one or two hidden layers, they can approximate [22] virtually any input-output map. They have been shown to approximate the performance of optimal statistical classifiers in difficult problem solving. Most neural network applications involve MLPs. The terms "Neural Network" (NN) and "Artificial Neural Network" (ANN) usually refer to a Multilayer Perceptron Network. It process the records one at a time, and "learn" by comparing their prediction of the record with the known actual record.

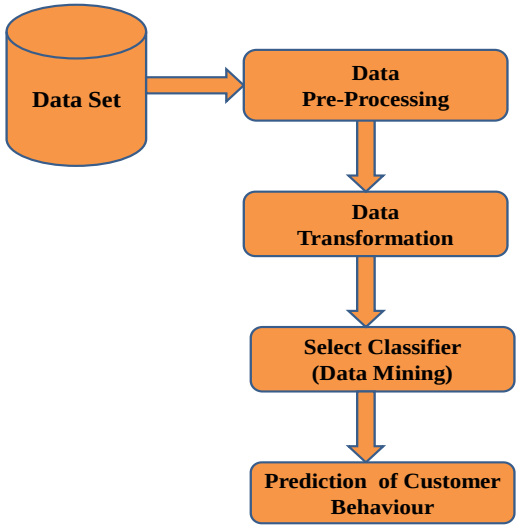


Figure 1: Steps involved in Predicting the Consumer Behaviour Analysis

4. Experiment Results

The data mining has novel methodologies for actual data testing and getting the result by implementing them. The researcher has to collect the data in a super market to get the customer views, product selection, associated product selection, buying capacity, an alternative selection etc.

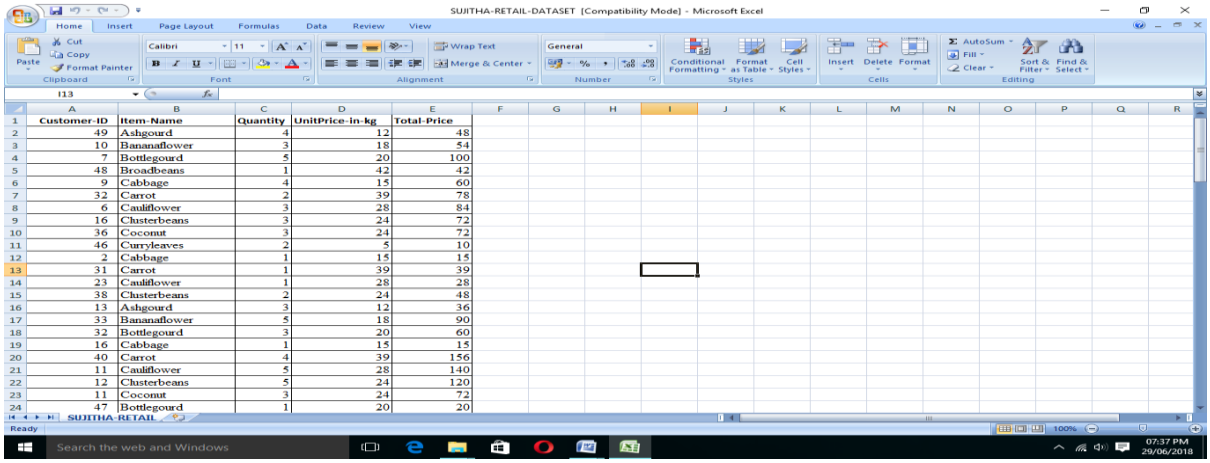
The data set is collected from various customers in a supermarket at sivagangai district. The analysis has included a sample of 100 data which is given Figure 2. The attributes are Customer-ID, Item-Name, Quantity, Unit Price-in-kg and Total-Price.

The data preprocessing is applied to extract the best attributes through cfs-subset evaluator and best-first search algorithm. The classification algorithms like J48 (C4.5), Random Tree and Multilayer Perceptron are applied on the data set. The objective is for predicting the Customer behaviour analysis in a super market. The algorithmic comparison by J48(C 4.5), Random Tree and Multilayer Perceptron interprets the liking and disliking of customer for various items. The outcome of the results is based on two decision tree algorithms'

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

J48 Pruned tree and Random. The interpretation is given in Figure 3 and 4. The outcome of the Multilayer Perceptron which is shown in Figure 5.

The comparisons of different classifiers are given in Table 1 and Table 2 and also the graph is shown Figure 6 and Figure 7. The accuracy for each algorithm is obtained from the Confusion matrix. The results of the algorithms are given in Table 3 and Figure 8.



Customer-ID	Item-Name	Quantity	UnitPrice-in-kg	Total-Price
49	Ashgourd	4	12	48
10	Bananaflower	3	18	54
7	Bottlegourd	5	20	100
48	Broadbeans	1	42	42
9	Cabbage	4	15	60
32	Carrot	2	39	78
6	Cauliflower	3	28	84
16	Clusterbeans	3	24	72
36	Coconut	3	24	72
46	Curryleaves	2	5	10
2	Cabbage	1	15	15
31	Carrot	1	39	39
23	Cauliflower	1	28	28
38	Clusterbeans	2	24	48
13	Ashgourd	3	12	36
33	Bananaflower	5	18	90
32	Bottlegourd	3	20	60
16	Cabbage	1	15	15
40	Carrot	4	39	156
11	Cauliflower	5	28	140
12	Clusterbeans	5	24	120
11	Coconut	3	24	72
47	Bottlegourd	1	20	20

Figure 2: Sample Data set for a Super market

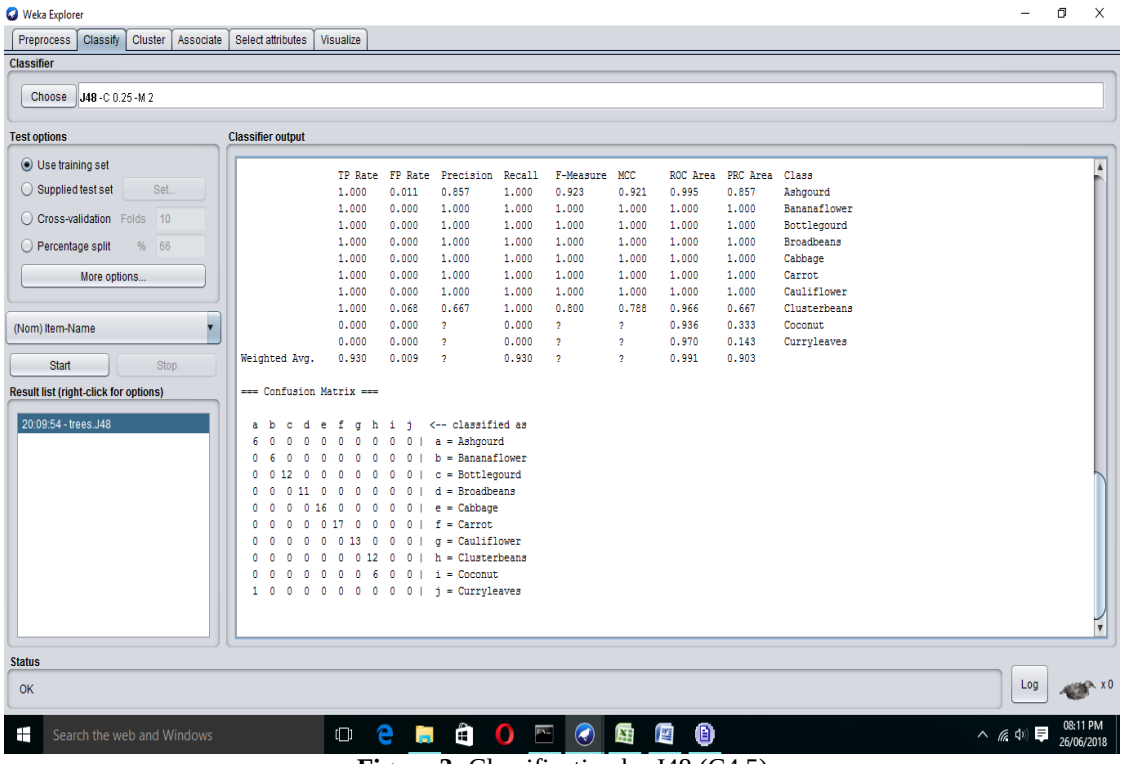


Figure 3: Classification by J48 (C4.5)

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

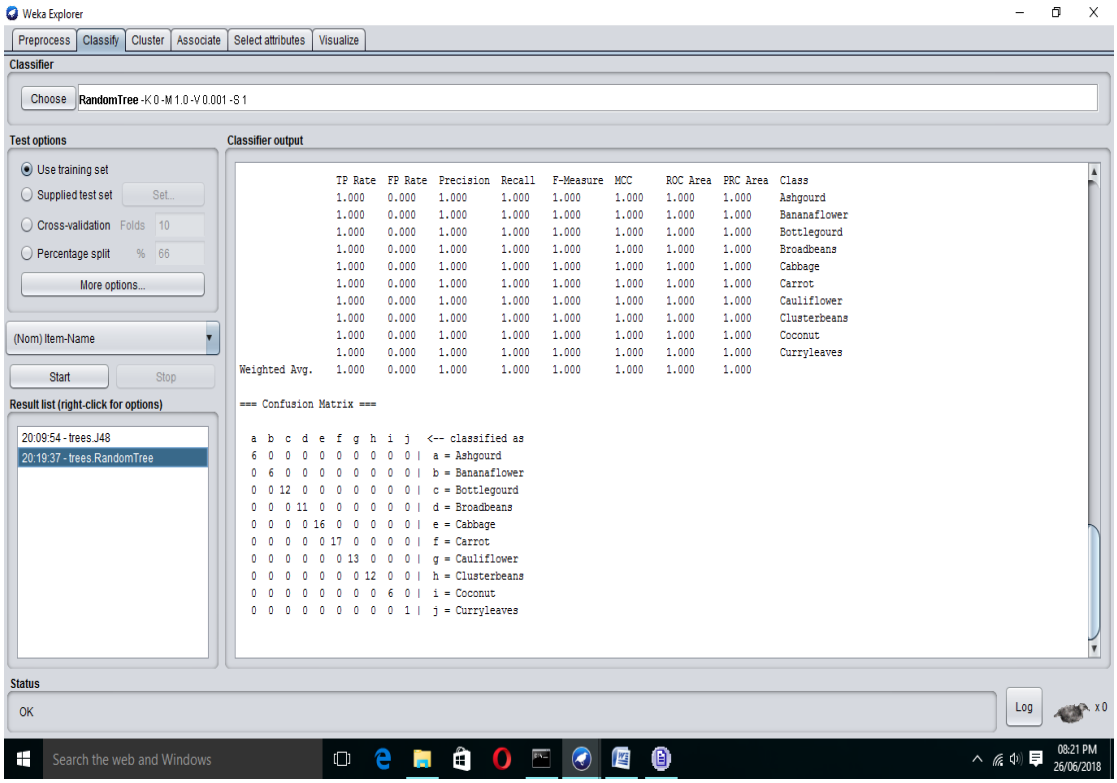


Figure 4 : Classification by Random Tree

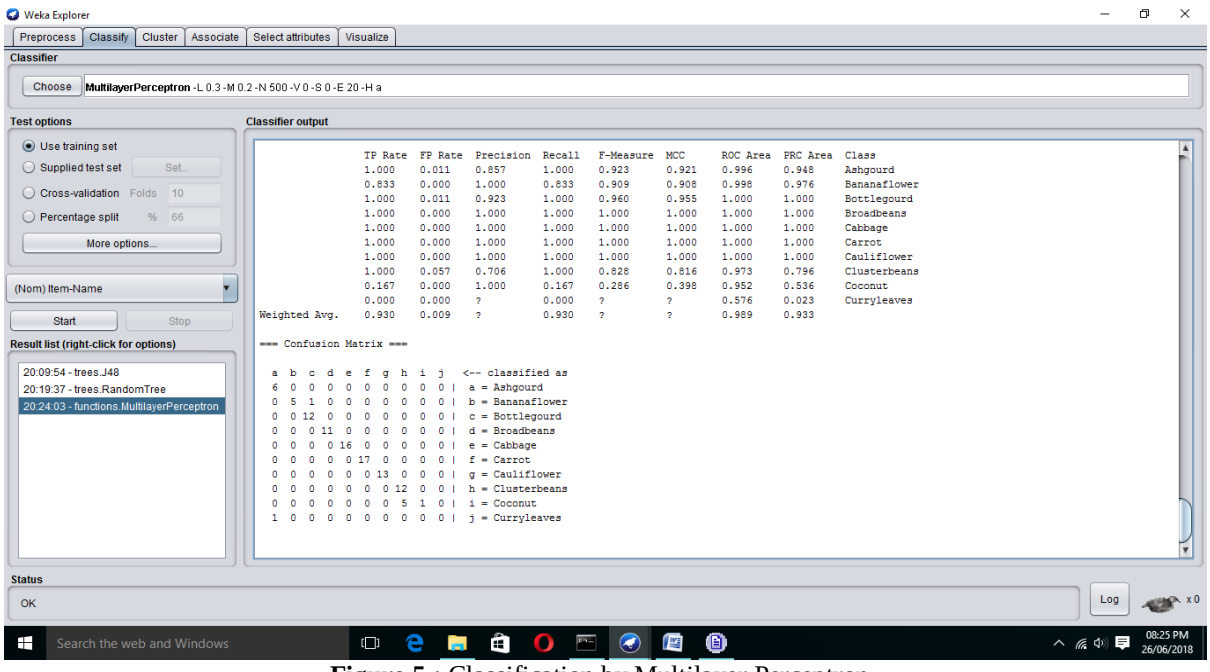


Figure 5 : Classification by Multilayer Perceptron

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

Table 1 : Classifications by Instances

Classifier-Name	**No-of-Instances-Correctly-Classified**	**No-of-Instances-Incorrectly-Classified**
J48 (C4.5)	93	7
Random Tree	100	0
Multilayer -Perceptron	93	7

Table 2 : Customer Liking of items by Classifier (TP Rate)

Classsifier	**Ashgourd**	**Bananaflower**	**Bottlegourd**	**Broadbeans**	**Cabbage**	**Carrot**	**Cauliflower**	**Cluster beans**	**Coconut**	**Curry leaves**
J48 (C4.5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000
Random Tree	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Multilayer -Perceptron	1.000	0.833	1.000	1.000	1.000	1.000	1.000	1.000	0.167	0.000

Table 3 : Error Accuracy

Classifier-Name	**MAE**	**RMSE**
J48 (C4.5)	0.0194	0.0986
Random Tree	0	0
Multilayer -Perceptron	0.0523	0.1267

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

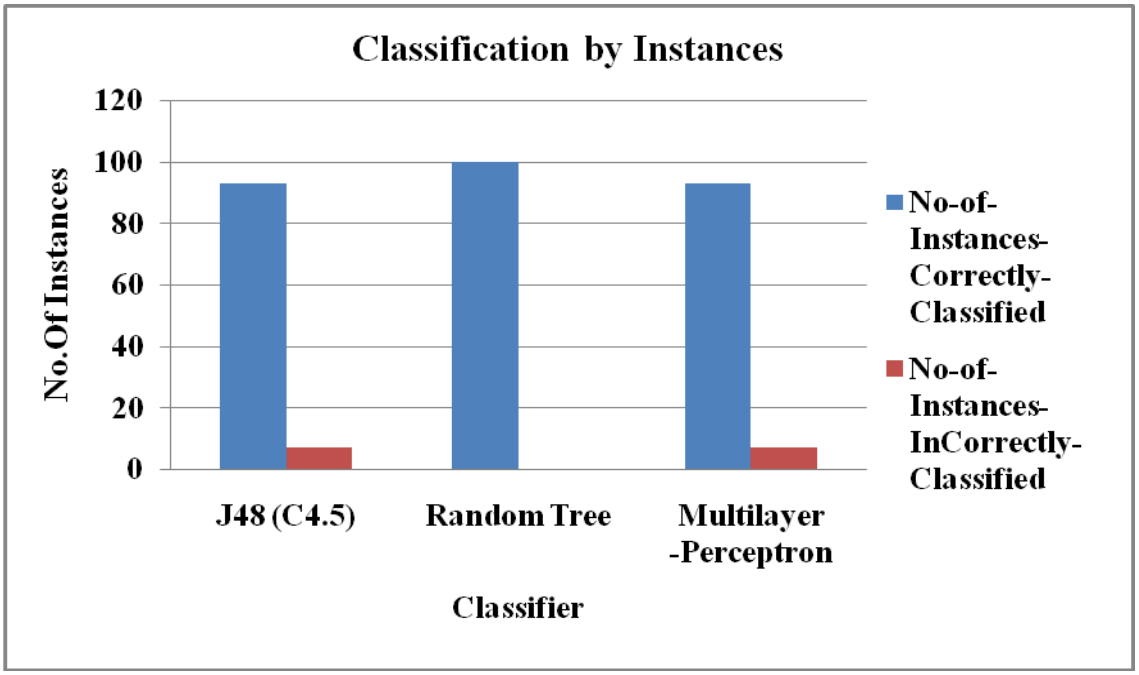


Figure 6: Classifications by Instances

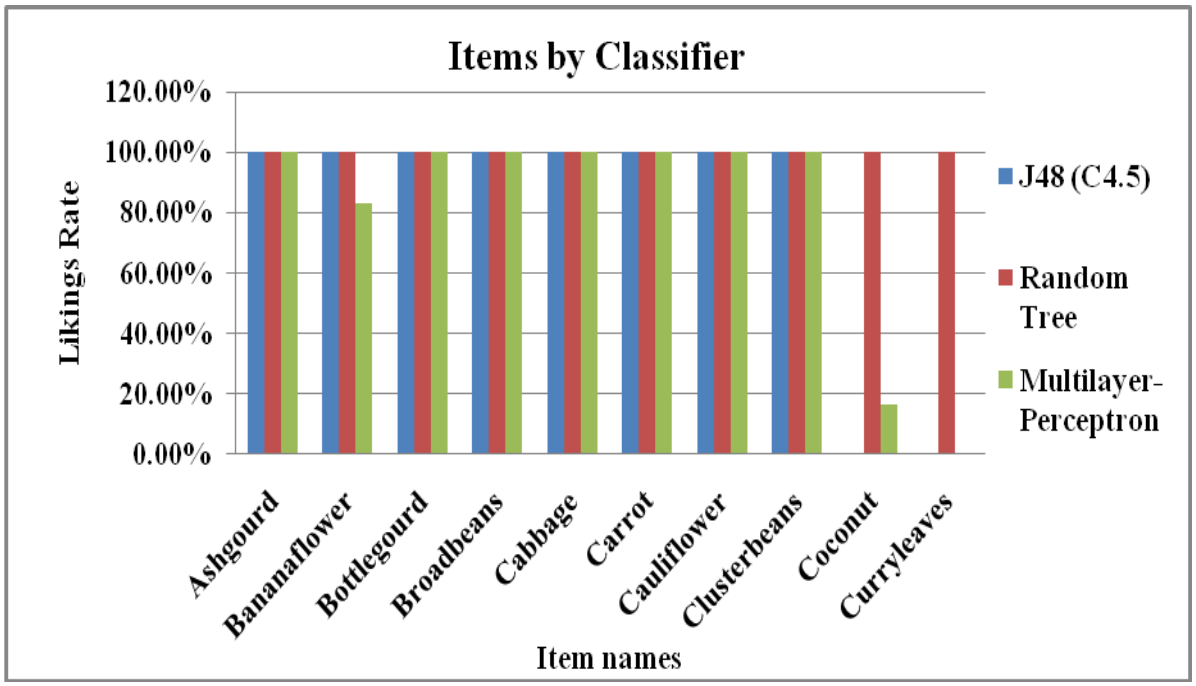


Figure 7 : Customer Liking of items by Classifier (TP Rate)

ISSN: 2321-8363
UGC Approved Journal
Impact Factor: 5.515

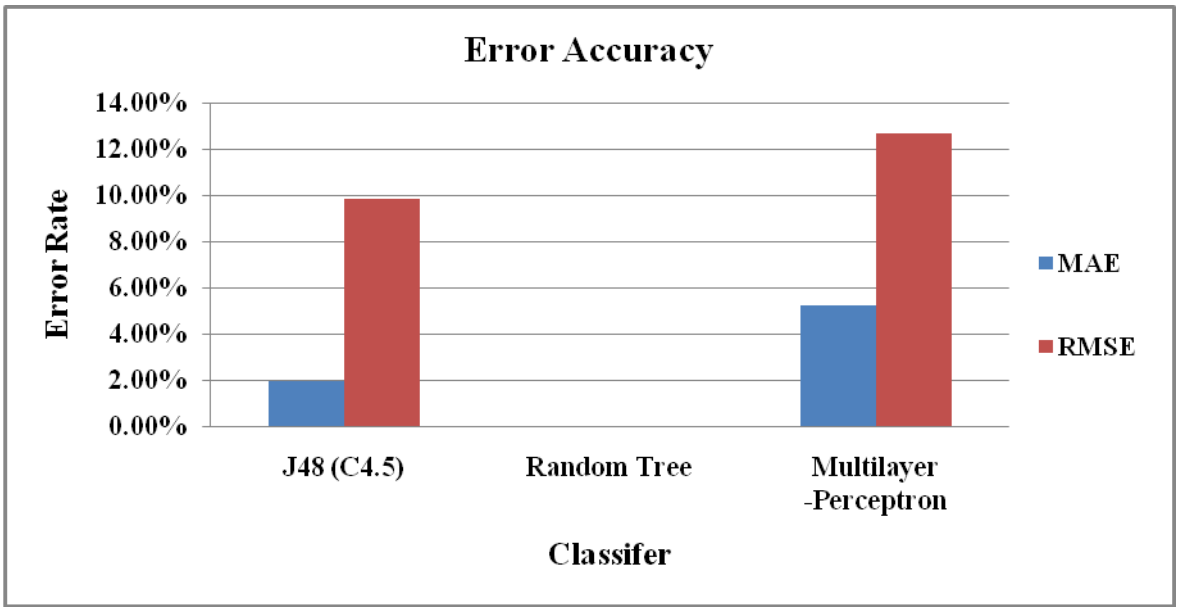


Figure 8: Error accuracy by Classifiers

5. Conclusion

In this paper, for analysis, there are 10 items has been selected from a top super market in sivagangai. The different sets of customers are selected and using the decision tree algorithm and the multilayer Perceptron is applied to analyze the buying rate (likings) of various customers. The three algorithms were applied to predict the True positive rate to understand the real liking of each customer. The random tree analysis gave good interpretation. The same gives 100 percent result comparing with other two algorithms. Similarly the error accuracy is best for the random tree classifier. Every super market can make an understand customer needs and the taste by seeing the graph in order to make a correct decision about the customers and also to improve their business**.**

REFERENCES

[1] Narasaiah .G.P and Jessica Saritha.S, 2013," Link analysis extension for mining relational databases ", *International Journal of Computing, Communications and Networking*, ISSN 2319-2720, Volume 2, No.4.

[2] Ayse cingoz and asuman Akdogan. A, 2013," Strategic flexibility, environmental dynamism, and innovation performance: an empirical study ", *Elsiver Ltd*, 9th International Strategic Management Conference, doi: 10.1016/j.sbspro.2013.10.528.

[3] Qasem A. Al-Radaideh, Ahmad Al Ananbeh and Emad M. Al-Shawakfa, August 2011," A classification model for predicting the suitable study track for school students", *IJRRAS*, Volume 8, Issue2.

[4] Anita Kumar, 2015," A study on surmising the performance of students using the data mining artistry", *International Journal of Recent Research in Mathematics Computer Science and Information Technology*, ISSN 2350-1022, Volume. 2, Issue 1.

[5] Nan-Chen Hsieh and Kuo-Chung Chu, 2009," Enhancing consumer behavior analysis by data mining techniques ", International Journal of Information and Management Sciences, Volume 20.

[6] Hasan Ziafat and Majid Shakeri, 2014," Using data mining techniques in customer segmentation ",*Hasan Ziafat Int. Journal of Engineering Research and Applications*,ISSN : 2248-9622, Volume. 4, Issue 9(Version 3).

[7] Monica LIA, 2015," Customer data analysis model using business intelligence tools in telecommunication companies ", *Database Systems Journal*, volume. VI, page no. 2.

[8] Dileep B. Desai and, Kulkarni .Dr. R.V, 2013," A review: application of data mining tools in crm for selected banks. ", *International Journal of Computer Science and Information Technologies*, Volume 4 (2).

[9] Srinivasa Kumar Dr.V,Renganathan Dr.R , VijayaBanu and Iyer Ramya ,2018," consumer buying pattern analysis using apriori association rule", *International Journal of Pure and Applied Mathematics*, Volume 119 .

[10] filipe pinto ,manuel filipe santos and alzira marques, 2009, "Database marketing intelligence supported by ontologies", *research gate* ,ISSN:1109-9526, volume 6, issue 3.

[11] Stuti Asthana, Amitkant Pandit and Anil Bhardwaj, 2017," Analysis of multiple hidden layer vs. accuracy in performance using back propagation neural network", *Indian Journal of Science and Technology*, Volume 10(4).

[12] Sujitha C, Umadevi Dr. B., "A Review On Customer Behaviour Analytics Through Business Intelligence", International Journal of Creative Research Thoughts, 2018, Volume 6, Issue 2 April 2018, ISSN: 2320-2882 IJCRT1812727.

[13] Umadevi, B., Sundar, D and Alli, Dr.P. 10th June 2014, "Novel Framework For The Portfolio Determination Using PSO Adopted Clustering Technique". *Journal of Theoretical and Applied Information Technology*. Volume. 64- No.

[14] Umadevi, B. Sundar, D and Alli, Dr.P. 26-28 Dec. 2013. "An Optimized Approach to Predict the Stock Market Behaviour and Investment Decision Making using Benchmark Algorithms for Naive Investors",*Computational Intelligence and Computing Research* (*ICCIC*). *IEEE International Conference*. Page(s):1 – 5978-1- 4799-1594-1.

[15] Umadevi, Dr. B and Dhanalakshmi, R. April 2017. "A Comprehensive Survey of Students Performance Using Various Data Mining Techniques". International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064. Volume 6 Issue 4.

[16] Naren Naga Pavan Prithvi and Tanneedi, 2016, "Churn prediction using big data analytics", *faculty of computing blekinge institute of technology*, issue no:37.

[17] Krishnan Ramanathan, Anil Jadhav . Dr., 2016, "Intelligent Decision Support Systems - A Tool For Human Resource Allocation In Information Technology Projects", *International Journal Of Computer Science And Information Technologies*, (IJCSIT) Volume. 7 (5) , 2356-2364.

[18] Cheolyong Park, "Simple Hypotheses Testing For The Number Of Trees In A Random Forest" , 17th March 2010, *Journal Of The Korean Data & Information Science Society* 21(2), 371–377.

[19] Mohsen Hayati and FarzadMozafari, 2012," Modeling and simulationof top-contact pentacene thin film transistor using artificial neutral networks", *Australian Journal of Basic and Applied Sciences*, ISSN 1991-8178.

[20] Gulshan Kumar and Krishna Kumar, 2012," The use of multi-objective genetic algorithm based approach to create ensemble of ann for intrusion detection", *International Journal of Intelligence Science*, Volume 2 ,page No.4.

[21] Gaurang Panchal, Amit Ganatra, Y P Kosta and Devyani Panchal, 2011 , "Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers",*international Journal of Computer Theory and Engineering*, ISSN: 1793-8201 , Volume 3, page No 2.

[22] Ranganath M S, Vipin and Mishra R.S, 2013," Neural network process modelling for turning of aluminium (6061) using cemented carbide inserts", *International Journal of Advance Research and Innovation*, ISSN 2347 – 3258, Volume 3.

Authors' Biography

C. SUJITHA 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.

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 27th Dec 2013 at VICKRAM College of Engineering and Technology.

