

Machine Learning Approach to Predict Potential Profitable Customers in CRM

Ruma Panda¹, DR. A. N. Nandakumar²

¹Assistant Professor, Dept. of C.S.E., Vemana I T, VTU, Bangalore, Karnataka, India

²Professor, City Engineering College, Bangalore, Karnataka, India

ABSTRACT: Customer Relationship Management helps to get customers by knowing their necessities, holding existing customers by fulfilling their necessities and drawing the attention of new customers by providing different marketing strategies. High value customers are performing a vital role to measure the effectiveness in CRM. The competition for High value customers is the central point of CRM.

Customer classification can help CRM to identify different type of customer for the growth of their organization. We have applied machine learning algorithms to classify the customers in CRM. Basically we have applied k-means clustering and Dynamic Rule based classification for this purpose. So the message can be delivered to the Potential Profitable customers based on their interest and the offer on products. It helps to keep the existing customers and draw new customers in this competitive market.

KEYWORDS: CRM, K-means Clustering, Twilio.

I. INTRODUCTION

In earlier days, generally sales team used to go out on the road or door to door to talk with customers about their products and discovering the important information. They used to maintain handwritten notes, laptops to keep all these information or they were trying to memorize those information. Afterwards sales team was using different platforms to contact with their manager or head of the company thorough phone, email and social media to discuss about customer reviews and to follow up their orders. Information can be missed or lost in that vast information without a common platform for customer interactions, communications and it leads to an unsatisfactory response to their customers. So there is a chance to lost the information, meetings are not monitored properly and can't priorities the customers. Even though they were used to collect data successfully, they faced with different types of problems. It can be challenging to extract intelligence and to find valuable customers. It is difficult to create the reports and it wastes valuable selling time. Managers can't give the proper support at the correct time to their team. So it can result a lots of mistakes and lack of responsibilities. In this scenario, CRM is very important to collect, manage, and use information which will decide whether you success or fail.

Now a days, many industries like Retail, Telecommunication, Insurance and Banking are using Customer Relationship Management (CRM) to wide their business. A CRM system can provide a perfect outline about customers of a particular organization. We can get to know the customer's previous history, the current status of their orders and customer service issues from the same place. CRM systems can keep the customer data from different channels like telephone, live chat, company's website, mail, marketing materials and social networks.

It also maintains customers' purchase history, buying preferences and personal information. Customer purchasing and interaction history can offer improved and quick customer service. It can also keep customer feedback and opinions about the organization from different social networks. Marketers can go through these reviews to improve their organization by implementing the new strategy which will help them to sale the product in a better way. It will give a clear picture of every opportunity and providing a proper way about sales. Enterprises should rely on customers to achieve profitable. It is needed to maintain valuable customers and satisfy customers by fulfilling their demands. The enterprises are needed to maximize the customer profits, extend the customer life cycle and improve the customer transfer costs in this competitive market. Customers are considered as essential part of the business process. But there are major challenges for retaining the customers and finding new customers to build high performance CRM models in the real-world application. So the customer is a central point of enterprise management strategy. Therefore, the goal of all enterprises is to maximize the profit. There are three possibilities to increase profits i) retaining existing customers by giving special offers, ii) obtaining new customers and iii) improving customer relationships.

II. LITERATURE REVIEW

RFM (Recency, Frequency, Monetary) model and association rule algorithm have been used to recognize valuable customers. It can also measure the similarity and difference which is depended on three rules, i.e. Emerging Patten Rule, Unexpected Change Rule, and Added/Perished Rule. It helps to find the current and hidden pattern of customers' behavioural changes. So management can identify possible variations of customer preference and delivers their product as earliest. They can expand their business by giving more preference to the customers and can retain them [1]. Customer data warehouse and mining are used to give the complete customers' information, find the valuable customers and identify the customer behavioural changes. Customers' behaviour has analysed to form the customers' profile properly under Internet and e-commerce environment. It helps to make more effective marketing strategy [2]. Naive Bayesian classification algorithm has been used to predict and classify the customer in Customer Relation Management for optimizing the business process. It helps the organization to identify marketing strategies and customer's pattern [3]. Electronic customers can be classified by applying a new Legendre wavelets neural network model. It provides the better accuracy of the transformation in electronic customer relationship management practically [5]. It can be

applicable in different industries such as banking, insurance industry, retail industry, manufacture industries, and so on. It is very difficult to classify the noisy and highly imbalanced data set in CRM. Different classifiers and new feature selection method are used to resolve this like SVM, Naïve Bayes and J48 classifiers. The maximum accuracy has been achieved in case of SVM and in J48 maximum precision and recall values have been achieved [6]. The aim of any competing businesses is to increase new customers and hold old ones. Mainly, every business should understand their customers' needs to provide the customer services and they can develop different marketing programs to the favour of their customers. Systematic customer segmentation is needed to understand this concept. A MATLAB program is used for k-means algorithm and a normalised dataset of 100 training examples is used to train the model and dataset is collected from a retail business [7]. A prediction model has been used to identify the customers who responded more for different offers based on their purchased history. Different classifications have been used to compare the efficiency of those techniques and identify the algorithm which is giving maximum accuracy for the existing data [8]. The vital role of the banking industry is credit risk. The prominent functions are gathering loan, credit card, investment, mortgage and others. Now a day's number of credit card users is growing enormously and a problem related to credit card default rate is increasing. Different algorithms like Logistic Regression, Decision tree, Random Forest tree are used for predicting credit default. Out of these, Random Forest provides higher accuracy [9]. Customer churn denotes to the occurrence of interruption of cooperation between customers and enterprises because of different marketing strategies. It can be analyzed by the customer's payment behaviour and business behaviour. It is possible to extract the attributes of the customer for a particular period of time. It helps to predict the probability of customer's loss in the future and their possible reasons. So it can find the possible lost customers very easily. Customer Churn Warning (CCW) algorithm is used to aware customers to churn [10]. In CRM, the essential phase is to convert the lead into actual customer. The fuzzy logic has been applied to show that how a lead can become a customer based on the collaboration between lead and business domain. Fuzzy logic is used to focus on the proper leads that have the potentiality to extend the upcoming sales [11]. Different clustering algorithms have been used to differentiate the customers and provide the different marketing policies accordingly. Hybrid clustering algorithm can perform well than individual algorithm in this scenario [12].

III. METHODOLOGY

We have applied machine learning algorithms to classify the customers properly. Basically we have used four steps for classifying the customer properly. It includes data normalization, k-means clustering, Dynamic Rule Based classification and messaging the upcoming offers to the Potential Profitable customers. We have collected the data for a mega retail business from Amazon. The dataset consists of 6 attributes with 2500 tuples.

A. Data normalization

It is needed to normalize the range of independent variables or features of data. It is needed to prepare the data. It helps to enhance the clustering performance. All training data are transformed to the range of -2 to +6. Normalization includes min-max, scaling and z-score. Here we have used z-score to normalize the data and then the normalized data is used for the algorithm.

$$X_{\text{norm}} = \frac{x - \mu_f}{\sigma_f} \quad \text{----- (1)}$$

The above equation is used to normalize the data using z-score where X_{norm} is the normalized value, x is the original value, μ_f is the mean value of features, σ_f is the standard deviation. Initially we have considered three centroids for three clusters. We have used orange color point to indicate the centroids and blue color square to represent customers in normalized form.

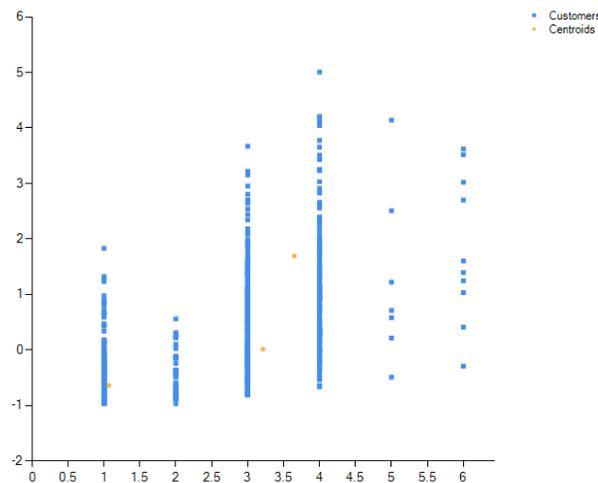


Fig. 1. Normalized data of k-Means algorithm

B. k-Means Clustering

According to [4], clustering algorithms generate clusters based on data similarity. We can use clustering in different areas like pattern recognition, medical science, banking etc. There are two types of clustering i) Hierarchical clustering ii) Partition clustering [6]. Hierarchical clustering algorithm forms the groups of similar objects depends on some hierarchies. There are two types of Hierarchical clustering i) top-down Hierarchical clustering ii) bottom-up Hierarchical clustering. More related objects will be grouped into the same cluster.

The top-down approach is also known as divisive clustering and the bottom-up approach is also called as agglomerative clustering. The various partitions can be created by Partition clustering algorithms and then evaluate them based on some criterion. k-Means algorithm is a famous

Partition clustering algorithm [4]. All data points will be positioned in any one of the k clusters which is chosen in prior. $\{x_1, x_2, \dots, x_n\}$ is a given set of n -dimensional training input vectors, k -Means algorithm divides the n training data sets into k sets cluster $S = \{S_1, S_2, \dots, S_k\}$, where $k \leq n$, such that the sum of squares is minimized within cluster. The k -Means algorithm works as follows:

1. Select the value for clusters k before running the algorithm.
2. Initialize the centroids for k clusters by shuffling the data points.
3. Nearest clusters are formed by assigning n data points.
4. Modify the clusters' centroids using the data points present in clusters.
5. Continue steps (iii) and (iv) until the centroids' position is becoming stable.

Basically three steps will be done in k -means algorithm. These are centroid initialization, assign data points and updating centroids. Minimum Square Euclidean distance is used to assign the data in the proper cluster. Based on that similar data will be grouped in the same cluster and dissimilar data will be grouped in different cluster. We have considered three clusters to classify the customers into High, Average and Low value customers.

C. Dynamic Rule Based classification

Always same data set will not be assigned to the same cluster like cluster1. Sometimes it will be assigned to cluster2 or cluster3. So it is very difficult to find a cluster where high value customer belongs. Because of that we have found high value cluster, mid value cluster and low value cluster dynamically by comparing the centroids of their clusters. Again we have applied dynamic rule based classification to classify each cluster. Actually we have considered these clusters to classify the customer more accurately based on last one year data. We have considered six types of customer. We have classified those three cluster customers into

- a. High value customer,
- b. Most Growth customer,
- c. Ordinary customer,
- d. Below Average customer,
- e. Negative Customer and
- f. New customer.

High value customers indicate highest purchased customer for last year. Most growth customers indicate that initially they were not purchasing the goods online but they started purchasing in last 3 months. Ordinary customers indicate that they purchase the goods infrequently and for fewer amounts. Negative customers never purchase any goods after creating their accounts. If any customer created his/her account in last three months then that customer will be considered as New customer without purchasing any item also. We can find the highest product item for last year by analyzing the purchased data history. Based on that company can change their business strategies to sale the other products or to sale the same product in a better way to grow their business. We can find the highest purchasing month and according to that company can give some attracting offers for that particular upcoming month in next year to attract more people. We can analyze the behavior of customers by analyzing their data.

D. Messaging the upcoming offers to the Potential Profitable customers

CRM companies are completely dependent on customers. It is very difficult to go through the complete customer data. Company is giving the priority to the Potential Profitable customers those are most likely to buy their products. Here High value customers and Most Growth customers are considered as Potential Profitable customers. Always CRM companies try to retain the existing customer and get new customers by giving some attracting offers. To do that, they send the upcoming offers through messages to the Potential Profitable customers to increase their business. So customers will be benefitted by getting the offers and they can recommend others also. We have used Twilio to send the messages for Potential Profitable customers.

IV. IMPLEMENTATION

We have applied k-Means algorithm after normalizing the data. k-Means algorithm groups the complete data into three clusters based on purchased history. We used different colours to represent different clusters where violet indicates High value customer, orange indicates Average customer and red indicates Low value customers in Fig. 2.

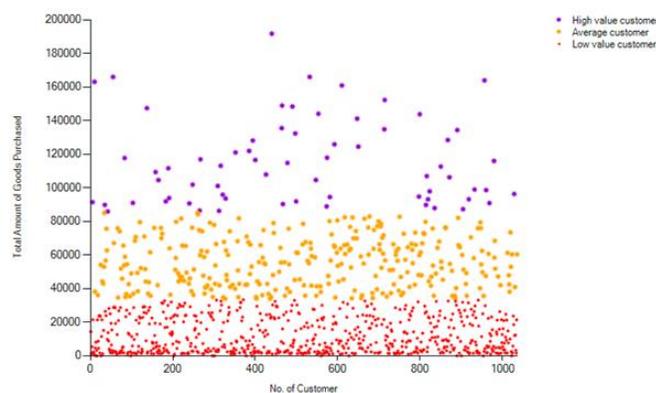


Fig. 2. Three clusters after k-Means algorithm

Now we used Dynamic Rule Based classification to classify three types of customers into six types of customers. We have plotted six different types of customers in Fig. 3 where pink represents High customer, orange represents Ordinary customers, green represents Below average customers, violet represents Most Growth customers, red represents New customers and blue represents Negative customers.

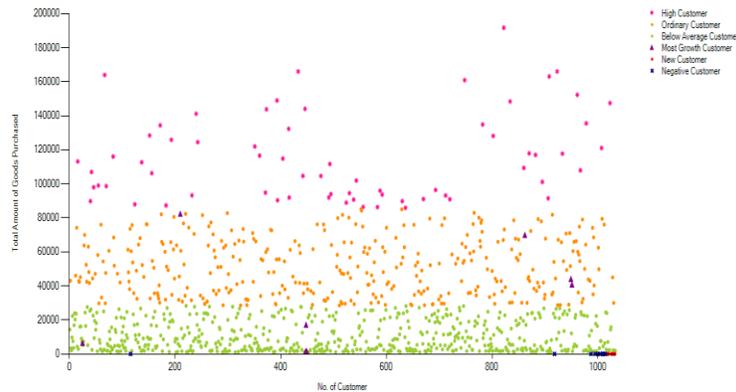


Fig. 3. Three clusters after k-Means algorithm

We can consider the different types of customers separately to analyze them in a better way. Fig. 4 presents High customer data for last one year. Similarly Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9 represent Most Growth customers, Ordinary customers, Below Average customers, New customers and Negative customers respectively. We can get to know about New customers and Negative customers from Fig. 8 and Fig. 9 like their user id and sign up date. Fig. 10 represents the total amount of goods purchased by customers and the no. of transactions made by them. So we will get to know maximum how many transactions happened and according to that, they can plan in a better way to increase the no. of transactions. For this, they should know that in which month, the maximum transactions happened. Fig. 11 will help to find out the no. of visitors based on month. Fig. 12 indicates amount of goods purchased per month. So it is possible to predict those valuable months when transactions are happening more. They can analyze the no. of customers for different products from Fig. 13. They can predict the valuable product to the customers and they can change their business strategies based on this like they can give some more attractive offers for other products to increase their sell. When the company will give some offers then automatically the messages will go to the high value, Most Growth and New customers. If the offers are good then they can recommend others also.

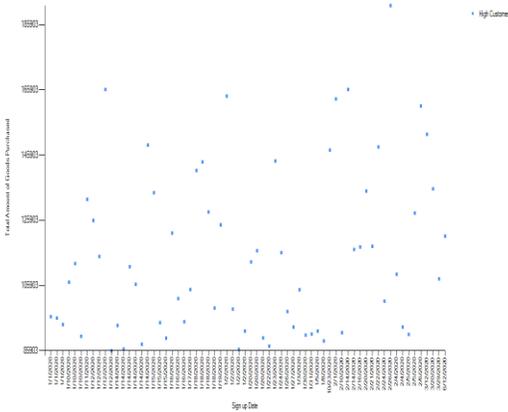


Fig. 4. High value customers

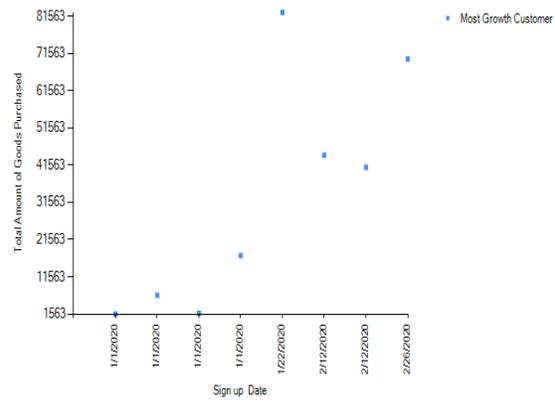


Fig. 5. Most Growth customers

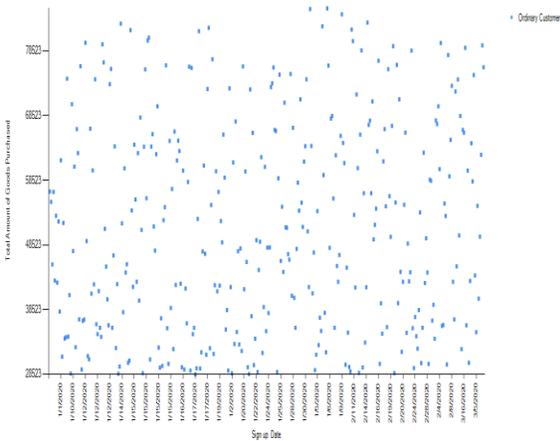


Fig. 6. Ordinary customers

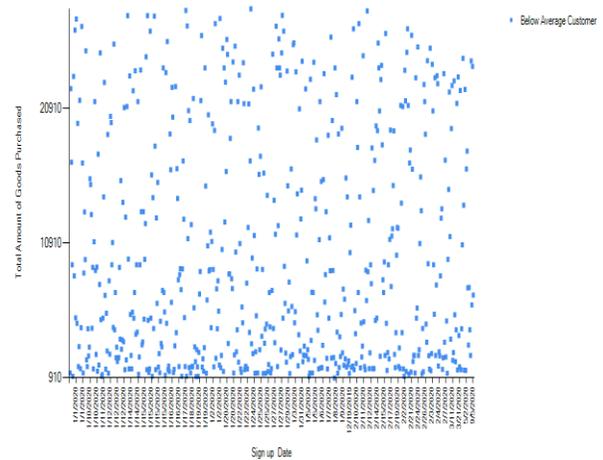


Fig. 7. Below Average customers

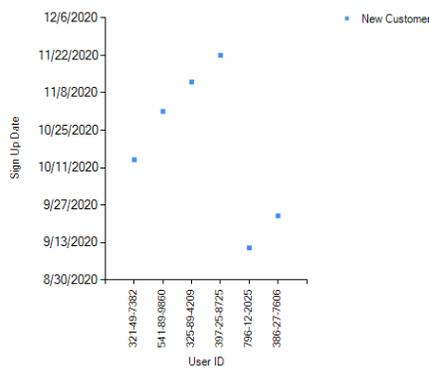


Fig. 8. New customers

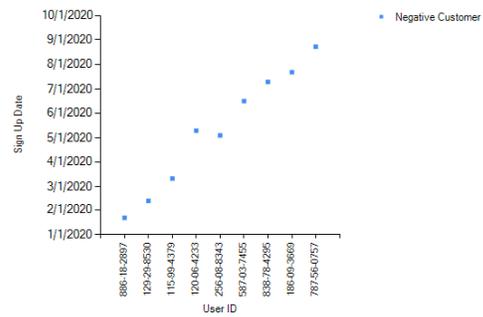


Fig. 9. Negative customers

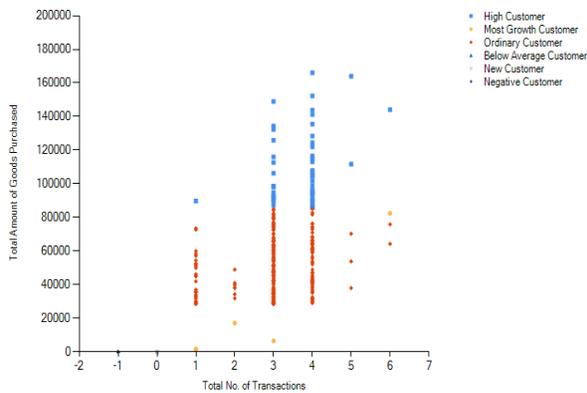


Fig. 10. Customer Details based on transaction

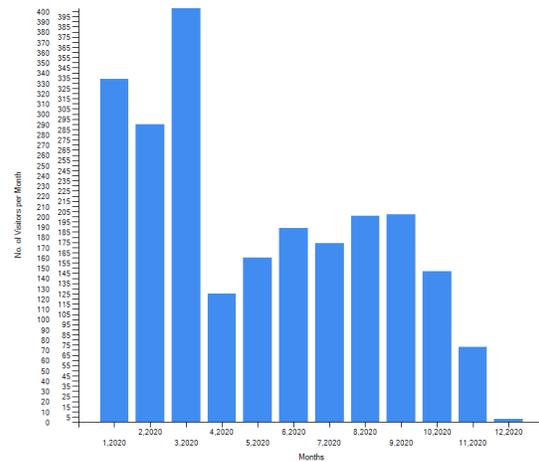


Fig. 11. No. of Customers visit per month

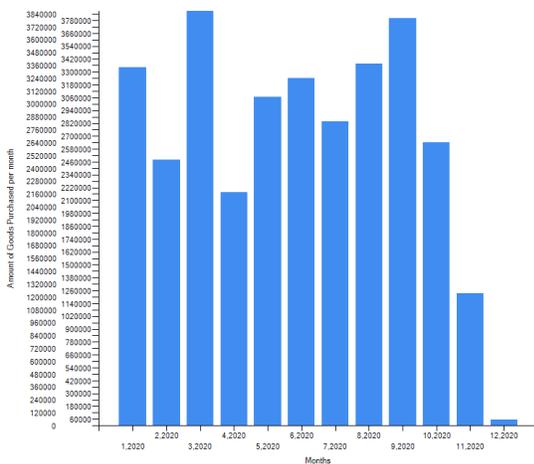


Fig. 12. Average Goods purchased per month

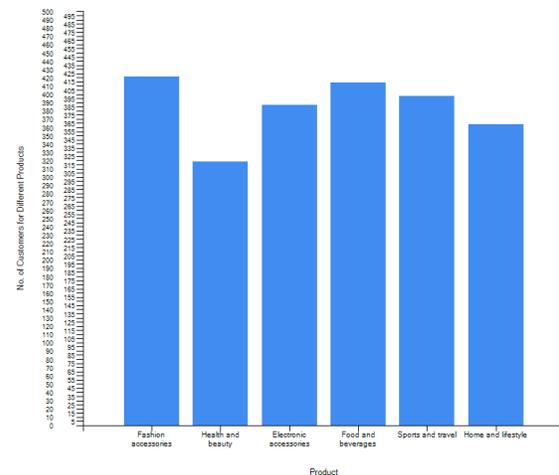


Fig. 13. No. of customers based on different products

VII. CONCLUSIONS

We have used C#.net and Twilio to implement this paper. We able to aggregate the customers properly by using these algorithms. It will help the company to avail the following benefits: can customize their market strategies which will be suitable for their customers, will be able to take business decision in case of risky situation such as credit relationship with its customers, can identify the valuable products and how to manage the demand and supply, can able to find potential profitable customers and defect customers, will try to find the association between products and customers which the business may not aware of and gathering additional market

research questions to provide guidelines of finding the solutions. Customers also will be benefitted in their future online marketing by receiving the messages from them.

REFERENCES

1. Ruey-Shun chen,Ruey-Chyi Wu and J.Y.Chen,"Data Mining Application in Customer Relationship Management of credit card Business", IEEE,2005
2. Che-Wei Chang , Chin-Tsai Lin , Lian-Qing Wang , "Mining the Text Information to Optimizing the Customer Relationship Management", Elsevier, 2009.
3. Gao Hua,"Customer Relationship Mngement Based on data Mining Technique", IEEE, 2011.
4. Vaishali R. Patell and Rupa G. Mehta, "Impact of Outlier Removal and Normalization Approach in Modified k-Means Clustering Algorithm", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011 ISSN (Online): 1694-0814.
5. L. Xinwu, "A new algorithm for electronic customer relationship management", Computing and Convergence Technology (ICCT), 7th International Conference, September 2012.
6. S.U.Natchiar,S.Baulkani"Customer Relationship Management Classification using Data Mining Techniques ",IEEE,2014.
7. Chinedu Pascal Ezenkwu, Simeon Ozuomba, Constance kalu, "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.10, 2015.
8. Tapan Kumar Das, "A customer classification prediction model based on machine learning techniques", International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), October, 2015.
9. Yashna Sayjadah, Ibrahim Akbar Tarigo Hashem, Faiz Alotaibi, Khairl Azhar Kasmiran, "Credit Card Default Prediction using Machine Learning Techniques", Fourth International Conference on Advances in computing, Communication & Automation(ICACCA), Subang Jaya, Malaysia, 2018, pp. 1-4.
10. Zuotian Wen , Jiali Yan , Liya Zhou , Yanxun Liu , Kebin Zhu , Zhu Guo , Yan Li, Fuquan Zhang, "Customer Churn Warning with Machine Learning", <https://www.researchgate.net/publication/329893308>, May 20019
11. Tanay Kulkarni, Purnima Mokadam, Jnanesh Bhat, Kailas Devadkar, "Potential Customer Classification in Customer Relationship Management Using Fuzzy Logic", International Conference on Innovative Data Communication Technologies and Application
12. pp 67-75, January, 2020.
13. Patel Monil, Patel Darshan, Rana Jecky, Chauhan Vimarsh, Prof. B. R. Bhatt, "Customer Segmentation using Machine Learning", International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429 Volume 8 Issue VI June 2020.

BIOGRAPHY

Ruma Panda received her M.Tech degree from West Bengal University of Technology and currently pursuing Ph.D. from VTU. She is associated with VEMANA IT as Assistant professor of CSE. Her area of interest includes Data Mining, Text mining, Design and Analysis of Algorithms, Formal Languages and Automata Theorem, Compiler Design.

Dr. A. N. Nandakumar is Professor in City Engineering College, Bangalore, Karnataka. He has done M.Tech in Computer Science & Engineering from Roorkee University in 1990. He received Ph.D. degree in Cloud Computing & Data Mining from Berhampur University in 2008. He published 35 papers in National Journal and 15 papers in International Journal.