Customer Attrition and Retention

Likhith D, Vamsi A, Rajashree Shettar

Abstract: In a competitive environment, organizations and firms are susceptible to customer attrition. Customer attrition and customer retention terms are widely spoken about. Customer retention which is quite the opposite of attrition is important for the company's sustainability in today's market. Many studies have come up with an attempt to find factors that influence customer retention. Firms have long desired to know who might end their relationship with them. Similarly, companies try to find how many existing customers did not return to purchase. Customer attrition is a problem that deals with clients and customers who are attrited from a particular brand or firm. In simple terminology it deals with the loss of profit associated with companies. This paper deals with the different ways to overcome the increasing attrition rate among customers. It also includes the implementation of customer segmentation using RFM model and K-means clustering. It also includes the prediction of customer retention using logistic regression.

Keywords: LVM (Lifetime Value Model), attrition, churn

I. INTRODUCTION

Today's world is a world growing rapidly in terms of technology and other aspects. With such an ever growing need, comes the growing spirit of competition amongst various brands and competitors. A hand full of brands are out there in the market to put out their product as the best causing a competitive market to arise in this sector. Customers are the asset of any business. It is very important to understand the customer base and their behaviour. In order to solve this we need to focus on customer segmentation. Customer segmentation helps in targeting customers in strategic approach to improve each user's experience which in turn aids in customer retention. After segmentation it becomes important to find out value of each customer, or in better words who amongst those customers that have attrited are the ones that hold a greater value to the brand in terms of profits or lifetime propositions. There are multiple attributive facts that can be used to check the value of a customer to a respective brand. One such feature is the CLV (Customer lifetime value) which is a measurement of how valuable a customer is to the business with an unlimited time span as opposed to just the first purchase. This metric helps you understand a reasonable cost per acquisition.

Revised Manuscript Received on May 30, 2020.
* Correspondence Author
Likhith D*, Bachelor of Engineering, Computer Science, Rashtreeya Vidyalaya College of Engineering. likhithd.cs16@rvce.edu.in,
Vamsi A , Bachelor of Engineering, Computer Science, Rashtreeya Vidyalaya College of Engineering. vamsikrishnasa.cs16@rvce.edu.in,
Dr. Rajashree Shettar , Professor, Dept. of Computer Science, RV College of Engineering, Bengaluru. rajashreeshettar@rvce.edu.in

Now comes the important concept of churn rate. When we talk in terms of a business context, "churn" is a terminology that refers both to the idea of customers' migration or can be to their loss of value. So, "churn rate" in simple terms, refers to the ratio of customers who terminate their relationship with the organization/brand, or, to the customers who still are active service receivers, but not as much their previous history suggests. Current organizations face a huge foregoing challenge, i.e. to be able to foresee those customers that have abandoned the brand in order to retain them on time, reducing this way costs and risks and gaining efficiency and competitivity. There are many advanced analytics tools and applications in the market, mainly focused to design and analyse the depth of the massive amount of data with regards to the firms, and hence make suitable predictions with regards to the information obtained from analysing and exploring the data at hand. Customer attrition is dealt with in a wide variety of industries ranging from e-markets, automobile industries, restaurant chains etc. Customer attrition is the loss of customers by a business. It is very important to monitor the attrition rate which helps in identifying strategies for improvement. The initial method is to perform data pre-processing. Exploratory data analysis is performed on a dataset in order to identify the main characteristics with the help of visual effects. The second step is to find if a customer will churn or not. Churning is predicted using the Pareto/NBD model. Pareto/NBD (negative binomial distribution) is the model used to predict the future activity of customers. It takes order history as input and also considers frequency and recency of orders. Customers who are likely to churn are differentiated based on their previous orders. Product categorization is done using Natural Language Processor. It provides specifications on the types of products bought by the customer. The third step is to perform clustering of customers based on product categorization. Semi-supervised learning used for customer segmentation in which RFM model scores the customers and segmentation is done with k-means clustering. Recency, frequency and monetary value are calculated for each customer and customers with lowest RFM score are to be considered. The clusters are analysed and clusters with high numbers of customers closely related are chosen. The previous steps are repeated for customers in a particular cluster. The module takes predicted churn customers and their respective cluster as input. This result is analysed and customers with low value to the business are not given any recommendation [4]. This paper attempts to convey various methods to decrease customer attrition in order to achieve growth in business. The paper proposes to solve the problem of a client at granular level. Customers are grouped to clusters based on their behaviour. RFM model scores each customer and k-means is used in clustering. Customer retention is predicted using regression.

This helps to improve the loyalty program and win back the existing customers.

II. RELATED WORK

Recency, Frequency and Monetary value are the analytical tools used to segment the customers into groups. The segmentation of customers is based on the RFM values. Recency is how recently did the customer purchase, Frequency is often the customer purchases during certain period of time, Monetary value is the average money spent on the purchases. Through these three multi layers, multi-view customer segmentation can be obtained [1].

RFM scoring is done based on the RFM values of each customer. Customers are divided into 4 quartiles, such that each customer will be assigned to one of the quartile in each dimension. RFM score is based on the quartile which customer belongs. RFM scores is used for segmentation of customers. The result of model can be used to predict the customer churn rate as well [1].

K-Means algorithm can be used to produce RFM scores based on the RFM values of each customer and finding the optimal K values for each of these three-layer data [1].

Churn analysis in simple terms is done with a lot of past evidential data. There are a vast variety of ways to do churn analysis which are most commonly supervised and unsupervised method [2].

Supervised methods, is basically a method that learns to classify data based on what it acclaims from given a specific training data set. The training data normally consists of a set of examples, for the purpose of training. In this method we have a particular output for every particular input. In this case the input object would be the rows of our data while the output value in this case would be a binomial value depicting whether the customer has churned or not. However in this particular case there is no mention of a particular attribute that will tell us who will churn. Hence this type of method is not completely optimal. Unsupervised methods: On the other hand, unsupervised methods refer to the problem by trying to find hidden structure in unlabelled data. We can use unsupervised methods to cluster our data set. This way the churners could become part of a separate cluster. Again, we can use methods which self-learn the data [2].

The customer's respective frequency is found out in this paper, and thus a depiction of their particular history can be formed, bucketed as either low frequency or high frequency [2].

By holding the two specific graphs side by side we can easily see that the line of the customer with higher frequency has inclined as time goes which means that the customer is using the service more over time. On the other hand, the line graph of the customer with lower frequency shows a declining line depicting lesser use of service over time. So our rule will correctly identify the first customer as not churner and the second one as churner [2].

Customer churn prediction is prediction of those customers who are likely to unsubscribe the service. Cost associated with customer acquisition is higher than the cost of customer retention. Linear regression and logistic regression are used to predict the churn rate. It is necessary to identify the problem of customer dissatisfaction and long-term relationship with customers. Customer dissatisfaction might lead to huge loss to the company. Long term relationships are an important factor in churning of customers. Bad relationships with customers will cause churn rates to increase. Increase in churning helps competitors to attract those customers. Churn rate can be found using both supervised method and unsupervised method. In this case, we have to build three models. The first model is the linear regression for the continuous data which gives the predicted probability of 76%.The second model is the logistic regression for discrete states which gives the predicted probability of 93%. The third model is the combination of the first two models using linear regression. The third model gives the predicted probability of 94% [3].

III. PROPOSED METHODOLOGY

The proposed method is to implement a RFM model for customer segmentation and logistic regression for customer retention. The initial method is to extract problem specific attributes from the raw data using SQL queries. These sets of attributes undergo data pre-processing. This method includes data cleaning, data transformation and data reduction. After the completion of data for pre-processing exploratory data analysis is performed to analyse the data set with visual methods. The next step is to build a RFM model where recency, frequency and monetary value need to be calculated. All the parameters should be considered while segmenting customers. After the calculation, RFM scoring is done on each customer based on their respective parameter values and RFM quartiles. Based on RFM scores, customers are clustered into groups and these groups are analysed repeatedly. Previous steps are repeated again until the cluster's centroid values do not change.

Once the customer segmentation is completed, customer lifetime value is predicted to know how valuable a customer is. Customer lifetime value is predicted using linear regression. Data set is divided into a training set and testing set. Fit the Linear regression model to the training set and then predict on the testing set. Hence the model will be able to predict the lifetime value of the new customer.

Once we have successfully segregated the respective dataset and also score each customer we move onto ranking these customers so that we know who to effectively target and thus can approach the marketing strategy better for those who have a high probability of being involved with the brand once again.

IV. RESULTS AND TABLES

The table shows the RFM values of customers. Recency, Frequency and Monetary Value is calculated for every customer. These values in each dimension are divided into 4 quartiles and scoring is done based the RFM values.

Customer key	Recency	Frequency	Monetary value
10004549	414	2	103.0018
10014326	619	1	65.96643
10058271	707	5	137.7568
10083104	1111	1	24.9385
10089955	418	1	0.4233

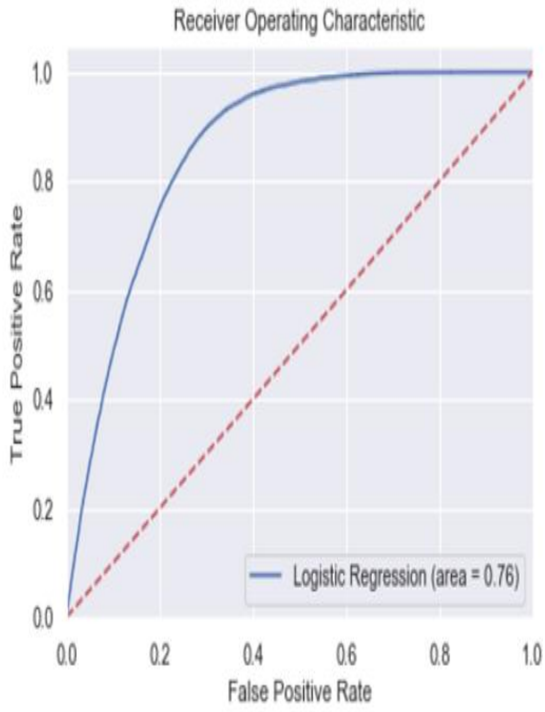
The table shows the RFM scoring based on the RFM values of each customer. Customers are divided into 4 quartiles, such that each customer will be assigned to one of the quartile in each dimension.

These scores of each dimension combine to form RFM score. The RFM score is used to Segmenting customers into groups. The customers with less RFM score are considered as best customers.

Customer key	Recency	Frequency	Monetary value	R quartile	F quartile	M quartile	RFM Class	RFM Score
10004549	414	2	103.0018	1	3	2	132	6
10014326	619	1	65.96643	2	4	2	242	8
10058271	707	5	137.7568	2	1	1	211	4
10083104	1111	1	24.9385	4	4	4	444	12
10089955	418	1	0.4233	1	4	4	144	9
10098559	1215	1	29.1326	4	4	4	444	12
10105292	811	1	14.1440	3	4	4	344	11
10106380	683	2	69.6921	2	3	2	232	7
10107178	586	1	68.6712	2	3	2	232	7
10128929	1330	2	36.0149	4	4	3	443	11
10155709	1097	2	49.4913	4	3	3	433	10

We obtain an ROC curve as shown above which shows an Area Under the Curve is portrayed is 0.76.We wish for the AUC to be as close to 1 as possible,as it results in hundred percent accuracy.
The ROC curve is plotted against the true positive rate and the false positive rate.

ROC - Curve



The confusion matrix shows the precision of the model along with the dataset size and the f1 score.
We observe it in both percentage and actual quantification. Logistic regression is applied for the model as seen.

Confusion Matrix

		Predicted	
		Y	N
Actual	Y	53%	11%
	N	11%	25%

		Predicted	
		Y	N
Actual	Y	21112	4,533
	N	4,238	9,773

Classification report

	Precision	Recall	F1-Score	Support
0	0.83	0.82	0.83	25,645
1	0.68	0.7	0.69	14,011

V. CONCLUSION

Customer attrition is a problem dealt by many industries and thus many strategies have come up as to dealing with this issue in order to solve customer retention i.e. in order to increase the retention rate. In this paper we basically depict the different methods and approaches that have been dealt with in order to deal with this issue and how optimal and beneficial these methods have turned out to be, thus implying a comparison amongst various such techniques.

REFERENCES

1. Afgan Fazri Handoko, Antoni Wibowo,"Three-Layer Data Clustering Model for Multi-View Customer Segmentation using K-Means",International Journal of Recent Technology and Engineering (IJRTE),Nov 2019
2. 2nd International Conference on Information Technology and Quantitative Management, ITQM 2014 Prediction of customer attrition of commercial banks based on SVM model
3. Pradeep B, Sushmitha Vishwanath Rao, Swati M Puranik and Akshay Hegde, "Analysis of Customer Churn prediction in Logistic Industry using Machine Learning," on International Journal of Scientific and Research Publications , 2007
4. Punya P Shetty, Varsha C M, Varsha D Vadone, Shalini Sarode, Pradeep Kumar D, "Customers Churn Prediction with Rfm Model and Building a Recommendation System using Semi-Supervised Learning in Retail Sector," International Journal of Recent Technology and Engineering (IJRTE),2019.
5. Oladipo K.A, Omotosho O.J, Adeduro O.A, "Predictive analysis for increased Loyalty and Customer retention in Telecommunication industry" International Journal of Computer Applications.

AUTHORS PROFILE

Likhith D is pursuing Bachelor of Engineering degree in the field of Computer Science at Rashtreeya Vidyalaya College of Engineering. His field of interests are Data Science and Machine Learning, Computer Graphics, Database Management System. He has done various projects such as Stock Prediction, Car Showroom Management System,Android Music App. likhithd.cs16@rvce.edu.in

Vamsi A is pursuing Bachelor of Engineering degree in the field of Computer Science at Rashtreeya Vidyalaya College of Engineering .His field of interests are Database Management System, Data Science and Machine Learning and Computer Networks. He has done projects such as Canteen Management System, Stock Prediction. vamsikrishnasa.cs16@rvce.edu.in

Dr. Rajashree Shettar is currently working as Professor, Dept. of Computer Science, RV College of Engineering, Bengaluru. Her research work focuses on "Knowledge Discovery in Semi-structured Data". She has around 45 publications in various International Journals and Conferences. She has authored a book on "Sequential Pattern Mining from Web Log Data: Concepts, Techniques and Applications of Web Usage Mining", published by LAMBERT Academic Publishing company, Germany and co-authored book chapters.