International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Understanding Customer Behaviour with Machine Learning

Rohan Bali, Satyajee Srivastava

Abstract: ***The study of customer behavior both in online and offline purchases plays a very important role for the seller. The aim of this study is to identify customers on various parameters and thus re-define policies based on the behavior of customers. This paper works on churn analytics for retaining customers, a market-based analysis for identifying the support and confidence among products and a recommendation system built on the IBCF approach. Churn Analytics helps the seller to answer about whether the customers are leaving there products or services. The goal of every seller is to maintain a low churn rate and thus have large margins and bigger profits. Further, performing a market-based analysis can be very fruitful for a supermart. This approach helps in organizing the items in a store in an efficient and scientific manner. This paper uses different machine learning algorithms techniques to conduct churn for the given data. It then calculates the accuracy and precision of each model using a confusion matrix. Confusion matrix thus helps us in selecting the best model to get more accurate results.This paper conducts the above analysis using the 'Apriori' algorithm. To conclude, a recommendation system is used to suggest customers products based on the history of their purchase or the similarities of that product with other products or other consumers. Thus, this study will help in understanding various aspects of customer behavior.***

Keywords: ***Analytics, Apriori, Churn, Customer Behaviour, IBCF***

I. INTRODUCTION

Companies that especially depend upon subscription-like revenue systems need churn analytics as a priority issue. Most of them have tons of data of each customer which can be fruitful in this process. Further, we need to understand the need for churn analytics. Basically, customer churn is a situation in which the customer finishes its relationship with the company or one of its subscription products. The fuel to every organization in this world is its customer base. Once, the customer base is affected that the organization takes a great deal of dip in its sales and profits as well.

Manuscript received on March 12, 2020.
Revised Manuscript received on March 25, 2020.
Manuscript published on March 30, 2020.

* Correspondence Author

Rohan Bali*, Student, Department of School of Computer Science and Engineering, Galgotias University , Greater Noida, India

Satyajee Srivastava, Assistant Professor**,** Department of School of Computer Science and Engineering, Galgotias University , Greater Noida, India

Adding to that it becomes more difficult for that company to add new customers as it is an expensive process and also the reputation affected by the customer that has churned affects the new customers that may purchase a product from them. The organization needs to cluster their customers into highly valued customers and not.

Highly valued customers can be defined as customers with expensive bills and a long relationship with the company. The company hence needs to take good care of them. These customers are extremely beneficial to the organization as they are not only high purchasers but also high supporters of their products thus can influence other customers to try out those products. The churn rate of this cluster of the customer should be the lowest. Customer Relationship Management or abbreviated as CRM plays an important role in maintaining the loyalty of the customers. Heavy investment is given for CRM for reducing the churn. The goal of the organization should be to reduce its churn rate. With the help of machine learning, this has become a more scientific and accurate way to deal with this problem. In this paper, we will use different techniques and tools/packages to solve this problem and also conclude with a comparative study. Along with churn analytics, this paper will reflect on market-based analysis and how it can be useful for an organization. Segmentation of different products in a super-mart can be fruitful. Making use of the packages available in R, we will conduct this research and find out the parameters 'support', 'confidence' and 'lift' to provide an efficient way of organizing products in a super-mart. The last part of this paper draws a comparative study between the two popular approaches to build a recommendation system i.e IBCF and UBCF. IBCF stands for Item-Based-Collaborative-Filtering and UBCF stands for User-Based-Collaborative-Filtering.

II. LITERATURE SURVEY

There have been various machine learning algorithms that can be used for churn analysis. The author in this paper discusses the implementation of logistics regression in conducting a churn analysis result[1]. Logistic regression is a classification technique in which all the data points are classified into binary outcomes. Further, after training the dataset, accuracy, and precision is checked using a confusion matrix. A confusion matrix has four quadrants, (namely TN-True Negative, TP- True Positive, FN-False Negative, FP-False Positive). These help in identifying the accuracy and precision of the model. A sigmoid function is used in logistic regression to identify the customers and classify them separately. The curve drawn by the logistic regression algorithm classifies the data into two binary outcomes. Further, the author in this paper uses a fuzzy technique to model churn[2].

Fuzzy logic can't be defined properly as it learns from the specific data and varies from each implementation where it has been applied. The clustering technique is also been used by one of the authors to form a different cluster based on the euclidean distance when plotted on a graph[3].

The approach used by these authors for identification os spam mail can also be used in churn analysis[4]. Decision tree classification was used by them. Information entropy helps in this algorithm when at each level data is divided and we gain more information about the data. Now, moving to market-based analysis, the authors in this paper provide an overview of association rule and its real-time applications[5].

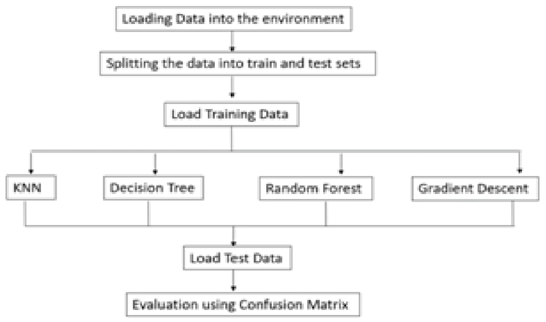
III. PROPOSED MODEL

We have tried to explain the different classification techniques that can be modeled on the data in order to determine the churn customers.

Each technique or algorithm used in this paper to explain the churn is evaluated based on a 'confusion matrix' and hence the efficiency and fit are determined. The modeling part starts with the simplest algorithm K-nearest neighbors (KNN) and then deals with the decision tree and finally we use the random forest as ensemble learning. We have chosen the set of algorithms in a way such that with each algorithm the flexibility increases. With the increase in flexibility the variance increases and bias decreases. The interpretability among different attributes of a given data is easy to understand with simple algorithms that are not very flexible. However, these algorithms generally don't lead us to very accurate result, thus we choose more flexible methods in order to determine the final result, but explaining the relationships b/w the variables become difficult. As a general rule, the aim is to predict an outcome we need more flexible methods but if the aim is to interpret the relationships among different variables we need less flexible methods.

(A) Mathematical Background

The K-Nearest Neighbour method using distances between the data points in order to classify the unknown data point. In this case, the customer needs to be classified into a churn group or not. The unseen data is plotted and k-neighbors are chosen for that point. Choosing the value of k is not an easy task. We can use the 'elbow method' to find k, but there are many more techniques for finding k. The distance is calculated using the Euclidean distance : $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$, where x and y are the do points plotted on a 2-D graph. Sometimes Manhatten distance : $\sum_{i=1}^{k}|x_i - y_i|$ is used. All the k-neighbors are measured from the unknown data point. The majority of the neighbors that are near that point makes the result that the new data point belongs to the majority of data points near it. For example, we plot a new customer in the graph. We will select k-nearest neighbors to it. If there are a majority number of data points that are churned as compared to non-churned, then the new customer is likely to get churned. Moving on from K-NN, we choose a decision tree model to deal with such a situation. A decision tree model simples split the training data at several points. Each splits gives rise to new information. This is known as an information entropy. The idea behind information entropy is that with every split made by the model, new information is supplied to the machine which helps it further to classify or predict the outcome of the model. The entropy results in information gain for the model. The mathematical formula of information entropy is given as $\boldsymbol{E(I) = \sum_{i=1}^{c} -p\, log_2\, p_i}$, where 'p' is the frequentist probability of an element or point in the data. The 'i' could be either positive or negative depending upon the sign convention we have used to classification groups. The entropy is the disorder in the data and the goal of every machine learning algorithm is to reduce the entropy. From the entropy, we get the information gain. The information gain is mathematically expressed as $\boldsymbol{IG(y,x) = E(y) - E(y|x)}$, where IG is the information gain from 'y' to 'x'. We just subtract the entropy of 'y' given 'x' from the entropy of 'y'. Greater the reduction in uncertainty, the higher is the information gain. The last step of a decision tree model, in this case, is to classify the customers in tow groups that which customers will churn and which customers will not churn. Moving to the random forest, as the name suggests, a random forest classifier, contains many uncorrelated decision trees that work as an ensemble. Each tree predicts the result as 'YES' or 'NO'. The majority of these predictions by the trees is given as the final outcome of the random forest. The key to a higher rate of accuracy in a random forest classifier is that the trees used in the modeling processes as an ensemble must be uncorrelated. The reason behind this is that these trees protect each other's errors, while some may go wrong, most of them will classify in the right way thus giving a rise in the accuracy of the model. The gradient descent model uses a regular linear model for classification. This method selects a random point and then the equation is differentiated in order to find the global minima. Then either depending upon 'learning rate' or choosing a point lower to the first random point this process is repeated till we find the global minima. This method will take the longest time to execute. It is an iterative program that optimizes our cost function and hence should be close to the most optimized and accurate result. Now, the second part of this research paper deals with the market-based analysis. Association rule learning is a machine learning concept that is used for finding interesting patterns and trends among variables in a dataset. The concepts that are used in association rule learning are; Support: It tells about the frequency of an itemset that appears in the data given. Let A and B be data items such that $\boldsymbol{x \to y}$, be an association rule on T purchases, $\boldsymbol{Support = \left(\frac{t \in T|_X \subseteq t}{|T|}\right)}$. Confidence: It represents how many times a given rule has been true. $\boldsymbol{Confidence = \frac{Support(x \cup y)}{Support(x)}}$. Lift: It is the ratio of the actual support to the expected support, given that X and Y are independent.



(B) System Workflow

The following flowchart describes the workflow of this paper.

STEP 1: Loading the data into the environment

STEP 2: Dealing with missing values.

STEP 3: Performing feature selection and subsetting important features.

STEP 4: Splitting the data into train and test sets.

STEP 5: Model the training data and form different classifiers.

STEP 6: Pass the test data to all trained classifiers.

STEP 7: Evaluate using a confusion matrix and measure accuracy and precision.

(C) Model Evaluation

This is the most important part of the research paper. The model evaluation is done using a confusion matrix. A confusion matrix sometimes also known as an error matrix, helps us in understanding the performance of a classification algorithm. A confusion matrix is a 2x2 matrix in which every row shows the predicted outcome and every column represents the actual results. Precision and recall are calculated from a confusion matrix. Precision is mathematically explained as $\boldsymbol{PR = \frac{TP}{TP+FP}}$, where TP is 'True Positive' and FP is 'False Positive'. The significance of precision is how accurately the model has predicted the right outcomes. Recall is mathematically described as $\boldsymbol{R = \frac{TP}{TP+FN}}$, where TP is 'True Positive' and FN is 'False Negative'**.** Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances, while recall (also known as sensitivity) is the fraction of the total amount of relevant instances that were actually retrieved.

IV. EXPERIMENTAL RESULTS

The experiment is carried out on a computer with the Windows 10 operating system, 16GB of RAM, and 1TB or harddrive. The following are the results of all the models that are represented in sperate confusion matrices.

Table1:K-NN-Confusion Matrix

	Actual	
Predict	No	Yes
No	2802	48
Yes	322	161

Table 2: Decision Tree Confusion Matrix

	Actual	
Predict	No	Yes
No	2825	25
Yes	154	329

Table 3: Random Forest Confusion Matrix

	Actual	
Predict	No	Yes
No	2815	35
Yes	122	361

Table 4: Gradient Descent Confusion Matrix

	Actual	
Predict	No	Yes
No	2812	38
Yes	239	244

From the above confusion matrices, we can calculate both recall and precision for all the models used in this problem. The following shows the accuracy of all the models when used on the test dataset.

Table 5: Precision and Recall of models(experimental results)

Model Name	Precision	Recall
K-NN	0.88	0.89
Decision Tree	0.91	0.92
Random Forest	0.94	0.94
Gradient Descent	0.95	0.95

From the above data, it is evident that gradient descent is the best-fit algorithm for this classification algorithm as suggested in the proposed model. However, it was also the model that took the longest time to execute so we might also need to consider the time-accuracy trade-off when dealing with this situation in real-time. The following shows the results of the market-based analysis when we worked on it using apriori algorithm.

Table 6: Relationships among products based on transaction history.

lhs	rhs	support	confidence	lift	count
{BANANAS}	{FLUID MILK WHITE ONLY}	0.053391	0.509099	2.336723	5903
{BANANAS}	{MAINSTREAM WHITE BREAD}	0.023082	0.220095	2.227176	2552
{BANANAS}	{DAIRY CASE 100% PURE JUICE - O}	0.019238	0.183441	3.012718	2127
{BANANAS}	{SOFT DRINKS 12/18&15PK CAN CAR}	0.018469	0.17611	1.901476	2042
{BANANAS}	{SHREDDED CHEESE}	0.017682	0.168607	2.313421	1955
{BANANAS}	{YOGURT NOT MULTI-PACKS}	0.017104	0.163088	3.470892	1891
{BANANAS}	{MAINSTREAM WHEAT/MULTIGRAIN BR}	0.014146	0.134886	2.787521	1564

The above shows us the relation between products i.e the relationship of 'bananas' with all the other major selling products in the supermart. We understand that those pairs of products that show high 'lift' must be placed apart from each other so that the customer needs to travel the whole store in order to buy them and in doing so he or she will purchase more than usual.

V. CONCLUSION

This paper shows us the different techniques that are used in order to identify the customers that are going to churn or not. Based on the above result we could see that Gradient Descent turned out to be the best working accurate solution for this problem as it is iterative in nature and its goal is to minimize the cost function which makes it an optimized or near optimized solution. The results from the decision tree and random forest were accurate too because of the information gain by entropy which takes place on every split made by a tree and hence giving more information to the machine. K-NN wasn't as accurate as of the other because it just depends on its neighbors and also the value of 'k' is difficult to determine in order to make this algorithm more optimized. Thus we can conclude that based on the above observations we can see that with the help of the confusion matrix the Gradient Descent model is the most suitable for this problem and hence must be deployed. Lastly, we conclude that based on the market analysis we concluded find confidence among several products which would be fruitful when the retailer is able to organize its store in an efficient manner for its increase in the sale. Thus, this paper draws some conclusions on how the customer behaviors can be judged and anticipated beforehand so that necessary changes can be made to retain the customer.

FUTURE WORK

In churn analysis, the use of bagging and boosting can also be done. Comparing the results of gradient descent and boosting techniques like XGBoost will be very close in terms of precision and recall. Hybrid models can also be built in order to increase the accuracy of the system. Then, their accuracies can be measured in the same manner as done in this paper. Apart from the churn and market-based analysis, we can build a recommendation system using an IBCF or UBCF approach for products in a supermart. This system would be beneficial for both sellers and buyers. Using the transactional data, an efficient recommendation system can be built and hence the customers with similar liking will be suggested products that are available in the store.

REFERENCES

1. Atul Sunil Choudhari and Dr. Manish Potey, **"**Predictive To Prescriptive Analysis For Customer Churn in Telecom Industry Using Hybrid Data Mining Techniques". Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),2018
2. K. H. Liao and H. E. Chueh, "Applying fuzzy data mining to telecom churn management," Intelligent Computing and Information Science, 2011, pp. 259-264.
3. C.-P. Wei and I. T. Chiu, "Turning telecommunications call details to churn Prediction: a data mining approach", Expert Syst. Appl., vol. 23, no. 2, 2002.
4. K. Binti Oseman, S. Binti M. Shukor, N. Abu Haris, and F. bin Abu Bakar, "DataMining in Churn Analysis Model for Telecommunication Industry", J. Stat. Model.Anal, vol.1,2010.
5. Achmad Bisri, "Hybrid Decision Tree and Logistic Regression Classifier for Email Spam Detection", Informatics Engineering Department Pamulang University Tangerang Selatan, Indonesia, 2016.
6. Pathan, Afrin & Palande, Snehal & Shende, Tejal & Patil, Ruchita & Vitthal, S & Gutte, & Devare, Manoj. (2019). A STUDY ON MARKET BASKET ANALYSIS AND ASSOCIATION MINING.

AUTHORS PROFILE

Rohan Bali is a data science enthusiast with a strong inclination towards problem-solving and propelling data-driven decisions. He is currently pursuing his Bachelor of Technology in Computer Science Engineering with a specialization in Data Analytics from Galgotias University. He has worked on multiple projects in the field of machine learning and deep learning. He is a keen learner. He also writes blogs on statistics and takes a keen interest in learning about new emerging technologies. He has two publications under his name in different books in the field of Big Data Analytics. He has also conducted full lab sessions in the university regarding machine learning and R programming. His field of research is big data, data analytics, and machine learning. E-mail: balirohan7@gmail.com

Dr. Satyajee Srivastava has over 17 years of teaching and industry experience. He received his Ph.D. in Computer Science and is an Artificial Intelligence developer, researcher, practitioner, and educator. He has Publication of 8 Books and 35 Articles in Journals of repute He is presently serving as Associate Professor at Galgotias University Greater Noida. He is an active machine learning researcher and regularly teaches courses and maintains resources for the data scientist. He is a member of the technical program committees for several technical conferences and editorial member of reputed journals. He has chaired, participated in and presented at conferences and seminars in India and across the world. Apart from academic pursuits, he has shouldered many administrative responsibilities in various capacities. His research has pioneered developments in ensemble learning, outlier detection and profile discovery Data Science Data Mining ,Big DATA and Green Computing. Email: drsatyajee@gmail.com