International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

Customer Relationship Management Using Machine Learning

Trisiladevi C. Nagavi[1], Manjari Ranjan[2], Anirudh D. Pai[3], Puneet Ahuja[4], Kavitha M. S.[5]

[1]*Assistant Professor, Dept. of Computer Science and Engineering, S. J. College of Engineering, JSS Science and Technology University, Mysore, Karnataka, India*
[2,3,4,5]*Undergraduate Student, Dept. of Computer Science and Engineering, S. J. College of Engineering, JSS Science and Technology University, Mysore, Karnataka, India*

Abstract – *At the present time, customers have shown more interest in the quality of service that the organization provides them. Many organizations provide services which cannot be highly distinguished. Such organizations need to maintain and increase their quality of service. They use customer relationship management to acquire new customers and maintain the relationship with the old customers. It also helps to increase customer retention for more profits.*

The customer relationship management systems adopt analysis techniques for understanding customer behaviours. These analysis systems use machine-learning models to analyse the customers' personal and behavioural data to give organizations a competitive advantage by increasing customer retention rate. When an organization tries to maintain customer relationships, different phases come into action like they have to identify customers based on geography or economic value, retain the customer and attract more customers. A customer relationship management system was successfully built that can perform analysis on different forms of customer data for the betterment of businesses.

***Key Words*: K-means, Gaussian NB, Gamma Gamma Model, Affinity Score**

1. INTRODUCTION

More than 30 percent of the world's population is a consumer of e-commerce. There are millions of products for sale provided by thousands of companies on the internet. In the last year, retail e-commerce sales came to 4.28 trillion U.S. dollars. These statistics indicate that we can conduct analysis for the improvement of both the consumers and the businesses.

Due to increased competitions among different organizations or vendors, quality of service plays a very important role. It is the way to maintain the customers and get new customers. Customer analysis systems implement machine-learning models to analyze customers' personal and behavioral data to give organizations an advantage by increasing the rate of customer retention.

There happen to be many such customers that remain with a company for a long time. There should be some way to know a customer's value to the business. In cases where the customer is very loyal, they can be rewarded with discount coupons or other kinds of rewards.

People spend hours on e-commerce services every time they want to buy something. Considering the amount of things we can get online, there should be some way to save the time spent on searching for products that are best for a particular user. Every major product release is followed by feedback by the users. The thousands of reviews made by the consumers need to be analyzed for getting an overview or an average feeling towards the product.

So, we have built modules to analyze customers that will help organizations to maintain customer relationships. For maintaining relationships, an organization needs to identify, retain and develop the customers.

1.1 Customer Identification

Its aim is to identify profitable customers and the ones who have higher chances to join the organization. On customer's personal and historical data, segmentation and clustering techniques can be performed to create groups of similar customers.

1.2 Customer Retention

For retaining customers, we build models to predict customer lifetime value. Customer lifetime value helps the organization to measure how valuable the customer is to them. Once companies know how valuable customers are to them, they can invest accordingly.

1.3 Customer Development

The primary aim of this phase is to increase the number of customer transactions for more profitability. Market basket analysis, feedback analysis and cross selling techniques are used.

2. LITERATURE REVIEW

Many researchers have analyzed customer behaviors using different approaches. Authors Adebola Orogun et al. [1] presented a paper on predicting consumer behaviors in the digital market based on machine learning approach. The aim here is to develop an association rule mining model to predict customer behavior using a typical online retail store for data collection and to extract important trends from this data. It was able to identify frequent itemsets, customer behavior features patterns and mining association rules between frequent purchase behaviors on an online store.

Sahar F. Sabbeh [2] has attempted to develop machine-learning based techniques for customer retention using a comparative study. Here, they tried to present a benchmark for the most widely used state of the arts for churn classification. The accuracy of the models was evaluated on the dataset of customers in a Telecom Company.

Pratik Thorat et al. [3] presented a paper on customer behavior analysis. They have identified customers related risks based on their purchases on e-commerce. Customer behaviors analysis is done by data mining and using machine learning KNN classification algorithm. Through the exhaustive experimentation, the customer who buys any unusual product or group/combination of unusual products is classified as High Risky customer and customer who buy usual products are classified as Low Risky Customer.

Vishwa Shrirame el al. [4] presented a paper on consumer behavior analytics using machine learning algorithms. This work aims to use data-driven processing tools such as data visualization, natural language processing, and machine learning models. This helps in understanding the demographics of organizations and in building recommending systems through collaborative filtering, neural networks and sentiment analysis.

3. METHODOLOGY

In this, a model was created where different phases of maintaining customer relationships are analyzed. For the identification phase, we build customer segmentation using K-means based on recency, frequency and monetary value. For retention of customer, we build models using gamma gamma to predict customer lifetime value of customers. So, they can invest accordingly. For developing customer relationships we build customer propensity, feedback analysis and recommending item models.

3.1 Customer Identification

3.1.1Customer Segmentation

To build this model, K-means clustering method was used. It is the algorithm that tries to minimize the distance of the points in a cluster with their centroid. Also, it is a centroid-based or a distance-based algorithm. Points are assigned to a cluster based on their distance calculated. Each cluster is associated with the centroid in K-means. The process is shown in figure 1.

Load the data

The dataset contains information related to all transactions that occurred for a UK-based and registered. It has 5,41,909 samples of data. . The data contains information on country, invoice no, stock code, description, invoice date and customer id. Load the necessary modules like pandas, numpy, seaborn, sklearn and matplot.

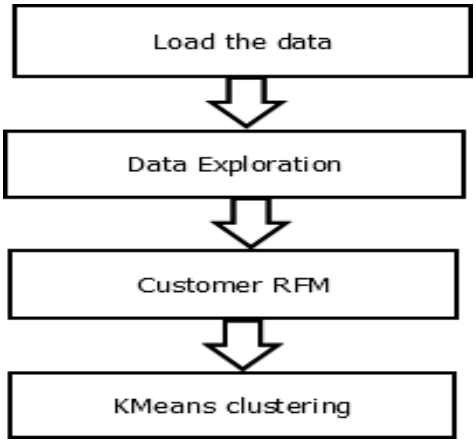


Figure 1: Flow Diagram of Customer Segmentation

Data Exploration

Remove duplicate and missing samples from the dataset. Perform statistical analysis on attributes unitprice, stock code. Then we have plotted a graph on daily sales, monthly sales and hourly sales. Same is shown in figure 2.

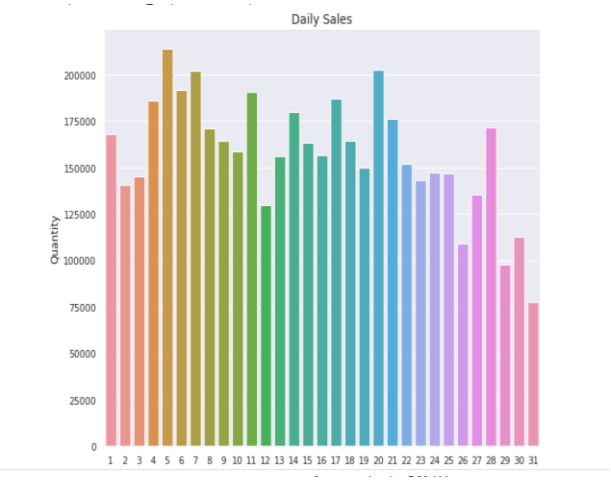


Figure 2: Daily Sales Graph

Customer RFM

Compute recency, frequency and monetary values for the given dataset. After calculating, make a new data frame consisting of customerid, recency, frequency and monetary values.

K-means Clustering

Clusters were formed based on customer's recency, frequency and monetary values. To find the number of clusters, we do elbow method to find optimal value of k.

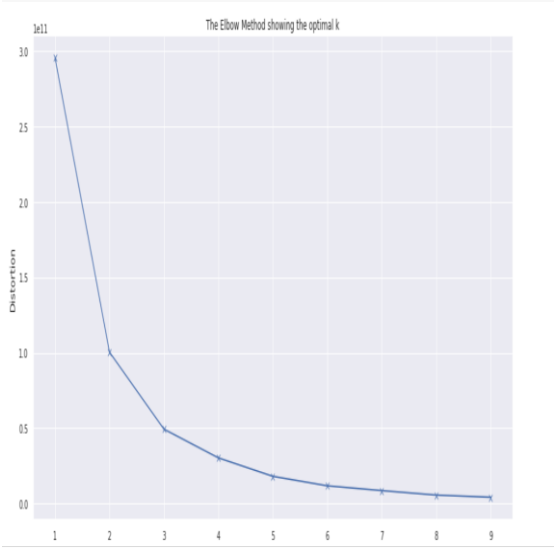


Figure 3: Elbow Method Graph

After looking at the graph shown in figure 3, k=4 was chosen as the optimal value because after that graph was linear. K-means analysis was performed using python libraries.

3.2 Customer Development

3.2.1 Customer Propensity

Dataset was downloaded from UCI repository. It has 455401 rows and 25 columns. The columns are UserID, basket add list, basket add detail, basket icon click, sort by, image picker, account page click, promo banner click, detail wishlist add etc. This dataset was trained on Gaussian NB model. Steps are indicated in figure 4.

Loading and Viewing Data

Load the necessary modules like pandas. The data contains information about the various links on the website that is clicked by the user during his browsing. This past data will be used to build the model. Load the training data and view the data types of fields in the dataset.

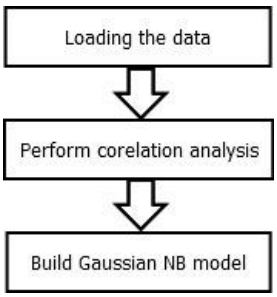


Figure 4: Flow Diagram Of Customer Propensity

Perform Correlation Analysis

Next step was to find whether there was any correlation between the user's individual website actions and an order, since we have all these fields in our data. With a predefined correlation function, correlation value was calculated. Attributes with less correlation value was dropped.

Build Gaussian NB Model

Split the data into training and testing data in the ratio of 70:30. Data was trained on Gaussian NB model to predict the propensity score. Naive Bayes classifier uses the Bayes Theorem. It predicts the membership probabilities for each group such as the probability that the given data points belong to particular groups.

3.2.2 Recommending Items

The CSV data file contains a list of items bought by users in the format of UserID and ItemID. Each item purchased by a user is tabulated as a pair of UserID and ItemID. The flow diagram of recommending items is depicted in figure 5.

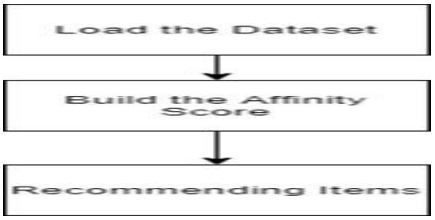


Figure 5: Flow Diagram of recommending items

Load the Dataset

The CSV file is read into the Data Frame by the read_csv function of the pandas package.

Build the Affinity Score

The affinity of all pairs of items was calculated. This was done by checking for how many item pairs were bought by how many customers per total number of customers. This gave a fair idea that if a person buys one item in the pair, then they have a chance of buying the other item too.

Recommending Items

For Item 1, it will predict item 2 with affinity score in decreasing order.

3.2.3 Feedback Analysis

An e-commerce website has millions of customers every day. Thousands of products are bought every hour and after a product is delivered, it is reviewed by the customer. The feedback made by the customer might be classified into many categories. Here we have considered 5 categories. They are bugs, complaints, comments, requests and meaningless.

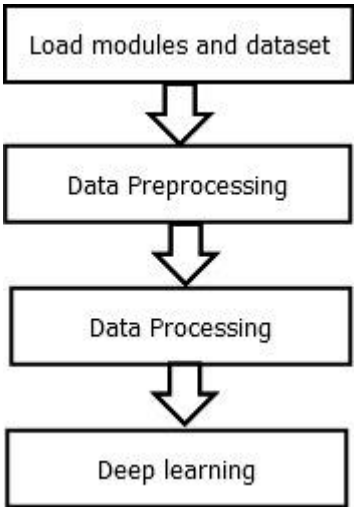


Figure 6: Flow Diagram Of Feedback Analysis

The dataset considered here has 5 text files for each category. In total these 5 files contain about 3200 reviews of different products from an e-commerce site. The text files contain lines of feedback. Each line in the text document represents one feedback submitted by a user. The flow diagram of feedback analysis is provided in figure 6.

Load Modules and dataset

The text files containing feedback from various customers are read into variables by converting each text file into a list of reviews. This is done by using the read line function and by specifying the path of the text files. These variables are then converted into a data frame containing two columns - text and category. The value of the column category is determined by the file that it was read from. It can be 'bug', 'complaints', 'comments', 'requests' or 'meaningless'.

Data preprocessing

The text data is put through preprocessing by removing hyperlinks, date, time and special characters. Then all stop words are removed from the text. They can be any words with low importance such as conjunctions.

Data processing

Convert the dataframe with preprocessed reviews to a matrix of TF-IDF features. Then run cross validation on different segments of the data frame. The average accuracy attained was 72%.

Deep learning

Using Keras some layers are added to the neural network and the model is run for 24 epochs. This is done to improve accuracy.

3.3 Customer Retention

3.3.1 Customer Lifetime Value

The dataset used for this model is downloaded from Kaggle. It is the same as the dataset referred to for the customer segmentation module. The flow diagram of customer lifetime value estimation is shown in figure 7.

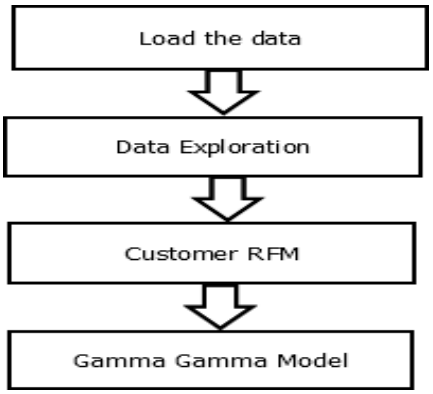


Figure 7: Flow Diagram For Customer lifetime value

Load Modules and dataset

Import modules Pandas, matplotlib, numpy, warnings and lifetimes. Pandas is needed for handling .csv files and converting them to a usable dataframe. Similarly, warnings is needed to filter out all the unnecessary warning messages. Matplotlib is used for showing visualizations using different available plots and lifetime packages for model development.

Data Exploration

Remove duplicate and missing samples from the dataset. Perform statistical analysis on attributes unitprice, stock code.

Customer RFM

Compute the values of recency, frequency and monetary for the given dataset. After calculating, make a new data frame consisting of customerid, recency, frequency and monetary values.

BG/NBD Model

In this model, built-in functions from the lifetime package are used to transform the data on transactions (one row per purchase) into summary data (frequency, recency, age and monetary).

Plotted the graph to visualize frequency/recency matrix and to predict if the customer is surely alive. The same is shown in figure 8. Yellow color part in the graph represents customers being surely alive.

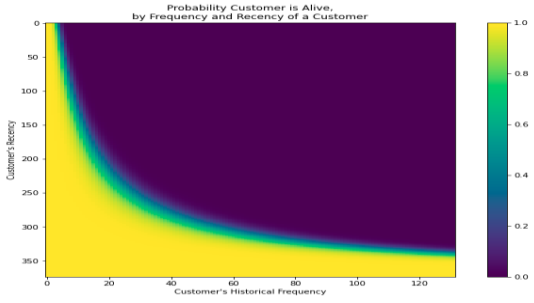


Figure 8: Graph of Customer being alive

Gamma Gamma Model

The model which we used to predict customer lifetime value for customers is called gamma gamma model. This assumes that there is no relation between the monetary value and purchase frequency. The model is trained to predict the conditional, expected average lifetime value of customers.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Customer Identification

4.1.1 Customer Segmentation

After k-means clustering technique, we divided our given data into four clusters as shown in figure 9.

Cluster 0 contains group of customers with low value of Recency, Frequency and Monetory.

Cluster 1 contains group of customers with high Monetory value.

Cluster 2 contains group of customers with high Frequency value.

Cluster 3 contains group of customers with moderate value of Recency, Frequency and Monetory.

```
[ ] # Kmeans with number of clusers = 4
    clusterer = KMeans(n_clusters=4, random_state=5)
    cluster_labels = clusterer.fit_predict(RFM)

    RFM['Cluster'] = cluster_labels
    RFM.groupby('Cluster').mean()
```

Cluster	Recency	Frequency	Monetory
0	93.502007	75.426919	1128.677111
1	2.666667	956.000000	241083.226667
2	6.866667	1508.466667	71538.928000
3	16.521008	476.268908	14304.523193

Figure 9 : Clusters based on RFM Values

4.2 Customer Development

4.2.1 Customer Propensity

Analyze the accuracy of the test data sample.

Accuracy was around 98% on testing the data on Gaussian NB model. The results are shown in figure 10.

A confusion matrix is plotted. It is used to describe the performance of the classification model.

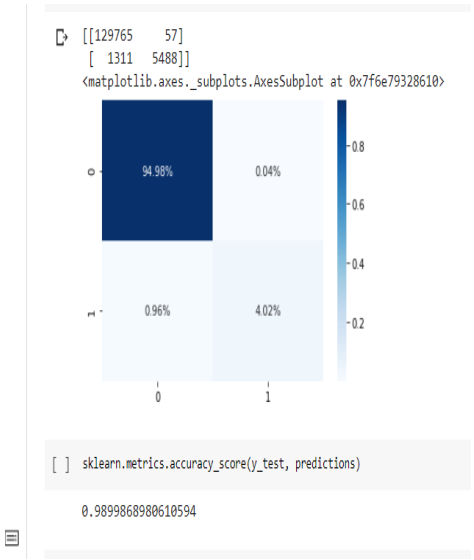


Figure 10: Accuracy of Gaussian NB Model

4.2.2 Feedback Analysis

Accuracy

Running the training for 24 epochs we ended up with 94.5% accuracy. During processing, there is a small amount of loss; however the gain in accuracy is high. The predicted feedback categories, accuracy and loss are projected in figures 11 and 12 respectively.

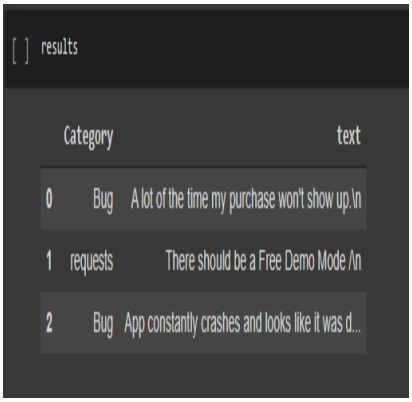


Figure 11: Predicting feedback categories

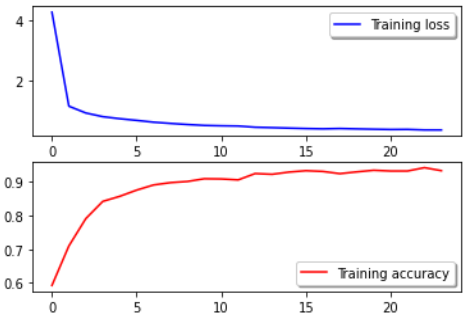


Figure 12: Feedback analysis accuracy and loss

4.2.3 Recommending Items

For item with id 5002, it recommended second item with their affinity score in decreasing order. The same is shown in figure 13.

```
searchItem=5002
recommendingitems=itemAffinity[itemAffinity.item1==searchItem]\
        [["item2","score"]]\
        .sort_values("score", ascending=[0])

print("Recommendations for item 5002 are\n", recommendingitems)
```

```
Recommendations for item 5002 are
      item2     score
1   5001.0  0.428571
14  5004.0  0.285714
12  5003.0  0.142857
16  5005.0  0.142857
18  5006.0  0.142857
20  5007.0  0.142857
```

Figure 13: Recommending Items

4.3 Customer Retention

4.3.1 Customer Lifetime Value

Gamma Gamma model is used to predict customer lifetime value for each customer. Same is pictorailly shown in figure 14.

```
data['CLV'] = round(ggf.customer_lifetime_value(
    bgf, #the model to use to predict the number of future transactions
    data['frequency'],
    data['recency'],
    data['T'],
    data['monetary_value'],
    time=12, # months
    discount_rate=0.01 # monthly discount rate ~ 12.7% annually
), 2)

data.drop(data.iloc[:, 0:6], inplace=True, axis=1)

data.sort_values(by='CLV', ascending=False).head(10).reset_index()
```

	CustomerID	CLV
0	14646.0	222128.93
1	18102.0	178895.33
2	16446.0	175531.47
3	17450.0	147476.62
4	14096.0	127589.20
5	14911.0	109442.13
6	12415.0	96290.23
7	14156.0	89440.33

Figure 14: CLV Prediction

5. CONCLUSION AND FUTURE WORK

Customer behavior analysis allows businesses to get a better perspective on customer lifetime, average purchase value, main audience segments and their needs. With all the required data and technology easily available, it leads to development and makes marketing much easier. One will know what niches to explore, how to appeal to potential customers and what functionalities to improve. Artificial intelligence and machine learning make the process much faster.

Customer identification is focused on through customer segmentation. We made clusters based on customer's recency, frequency and monetary values. For customer retention, we predicted clv value for each customer. In customer development phase, the main objective of this phase was to increase the number of customers. In the proposed system, we have recommended items, calculated propensity and did feedback analysis. For all the models different datasets are used. So, our future goal is to build all the models from one datasets and integrate the models.

ACKNOWLEDGEMENT

The authors would like to thank the management of S. J. College of Engineering, JSS Science and Technology University, Mysore, India.

REFERENCES

[1] Adebola Orogun, Bukola Onyekwelu. Predicting Consumer Behaviour in Digital Market: A Machine Learning Approach, International Journal of Innovative Research in Science, Engineering and Technology , Volume 8, Issue 8, August 2019.

[2] Sahar F. Sabbeh. Machine-Learning Techniques for Customer Retention: A Comparative Study, International Journal of Advanced Computer Science and Applications(IJACSA), Volume 9, Issue 2, 2018.

[3] Pratik Thorat, Dhanamma Jagli. Customer Behavior Analysis: Identifying risky customers based on their purchased product on e-commerce. International Research Journal of Engineering and Technology, Volume 7, Issue 9, September 2020.

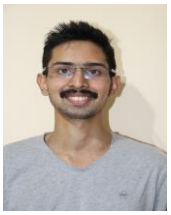
[4] Vishwa Shrirame, Juyee Sabade, Hitesh Soneta, M Vijayalakshmi. Consumer Behavior Analytics using Machine Learning Algorithms, 2020 IEEE International Conference on Electronics, Computing and Communication Technologies.

BIOGRAPHIES

Dr. Trisiladevi C. Nagavi is an assistant professor in the Department of CS&E. She graduated from Karnataka University, Dharwad. She obtained her Master's and Doctoral degree from Visvesvaraya Technological University, Belgaum. She has an expertise in the area of Audio, Music, Speech and Image Signal Processing, Information Retrieval and Machine Learning. Her research outcome resulted in an Android App to play favorite tunes. It is designed to retrieve songs by listening to user hum based on music melody representation models. She is an Active member of IEEE India Special Interest Group on Communications Disability and worked on Real Time Text to Speech Conversion and Translation System in Collaboration with All India

Institute of Speech and Hearing (AIISH) Mysore and IEEE Standards Association.

Manjari Ranjan is a final year undergraduate student pursuing Computer science and engineering from JSS Science and Technology University Mysore Karnataka. Her interests include machine learning and cloud computing.



Anirudh D. Pai is a final year undergraduate student pursuing Computer science and engineering from JSS Science and Technology University Mysore Karnataka. His interests include image processing and data analytics.

Puneet Ahuja is a final year undergraduate student pursuing Computer science and engineering from JSS Science and Technology University Mysore Karnataka. His interests include web development and networking.

Kavitha M. S. is a final year undergraduate student pursuing Computer science and engineering from JSS Science and Technology University Mysore Karnataka. Her interests include web development and cloud computing.