LRFM model for customer purchase behaviour using K-Means algorithm

C Jamunadevi1*, S Tamil Selvan2, M Govindarajan3, C Saravanan4 and B R Janaki raman5

1Assistant Professor, Department of CT-PG, Kongu Engineering College, Perundurai, Erode, India.
2Assistant Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College, Erode, India.
3Associate Professor, Department of Computer Applications, Velalar College of Engineering and Technology, Erode, India.
4Assistant Professor, Department of Information Technology, Velalar College of Engineering and Technology, Erode, India.
5UG Scholar, Department of Computer Technology-PG, Kongu Engineering College, Perundurai, Erode, India.

*Email: jamu83@gmail.com

ABSTRACT - The COVID-19 pandemics have a major collision on every aspect of life, including how people shop for their requirements. As the pandemic has reshaped life as we know, it’s also initiated many trends – but the biggest of these trends may be online shopping. The shift toward online shopping was happening before the pandemic, but according to new statistics from IBM, the COVID-19 has accelerated consumers shift toward online shopping by 5 years. The chief idea of the article is to inspect if the situation is approaching people to purchase things online and the continuation of shopping things online even after the end of pandemic. The information for the article has been gathered by circulating the survey on social networks. The questionnaire is comprised of 12 different questions, and 615 people responded to it. This work is based on LRFM (Length, Recency, Frequency, and Monetary) replica and separation of data based on the questionnaire using K-Means algorithm. Silhouette analysis helps to decide the extent of division among clusters. The results of the survey has a termination that people are fond of purchasing products online through the lockdown and people too agreed that the rate of online shopping will increase in the future when this pandemic is over.

1. Introduction
The COVID-19 pandemic is one of the highest worldwide healthiness crises of the recent time. The ultimate confront we have never seen after World War II. COVID-19 was first initiated in a city known as Wuhan of China in December 2019. Till now (25th September 2020) approximately 9, 83,000 people were died due to COVID 19 and nearly 32.3 million people were affected by it. COVID 19 has multiply to the whole world. The WHO acknowledged coronavirus as an Emergency of worldwide anxiety on January 30, 2020. This pandemic is to a great extent than a healthiness crisis; it’s also an incomparable social-financial crisis. Stressing each one of the countries it touches; it has the prospective to craft shocking economic, political, and social possessions that will depart profound and venerable scars. Coronavirus has hastily pretentious our routine living, businesses, disrupted world trade, and actions. The COVID-19 outburst has enforced many businesses to shut, foremost to unparalleled trouble of commerce in many industry sectors. Brands and retailers face many short-
range challenges, such as those related to health and safety, the workforce, the supply chain, cash flow, sales, marketing, and consumer demand. On the other hand, successfully solving these challenges will assure a promising future. This is because once we get through a pandemic, we will come out in a very different world compared to the one before the outbreak. Many markets, especially online shopping, are seeing unprecedented growth. This paper intends to examine the growth of online shopping in the future days by taking a survey from the people through the Google form questionnaire and an analysis is made through the K-means Clustering algorithm.

2. Literature survey

This paper[1] describes about the customer segmentation based on their buying history and the silhouette value is calculated based on the recency, frequency and monetary value using the silhouette value customer behaviour is predicted.

In this paper[2] the authors proposed the recent trends and the growth of e-commerce during the pandemic situation worldwide. They have also described about the top sales of e-commerce during the pandemic situation and discussed the way which e-commerce provide a way for the people to satisfy their needs.

This paper[3] proposes the statistical and percentage analysis of people doing online shopping and how corona virus is approaching people to perform online shopping rather than the physical shopping. The acceptance of people to purchase things online and their comfort while doing online shopping is discussed in brief using the graph.

This paper[4] Comprises a special data mining framework known CRM to predict the customer behaviour. They were classified into two for better performance. This involves the best model to bring the prediction to a higher rate is MLPNN and an accuracy of 88.63% is achieved.

In this paper[5] the authors delivered a way for customer segmentation with some benefits and calculated the performance of clustering approaches. They provide a result by analysing different clustering approaches and delusion of each approach.

This paper[6] provides an endeavour to comprehend the unrest impact on the economy, we sum up the impact of COVID-19 on singular parts of the world economy, zeroing in on essential areas which incorporate businesses engaged with the creation of completed items and tertiary areas including all help arrangement ventures.

The paper[7] the creators made a study of 200 client in a gem dealer and dissect the client conduct dependent on their decisions and their needs to purchase a gem.

The paper[8] Web based business has started better approach for appropriation channel and internet shopping inclination to expand the requirements and needs of client and the progress in data innovation.

3. Proposed methodology

3.1 Steps to be implemented

Step 1: Framing Questionnaires’ and collecting the responses

To know the rate of increase of online shopping trend, we have prepared a questionnaire which comprises of 12 questions. From these questions the online shopping activities of the people can be analysed before and after the pandemic lockdown. Through Google form we have collected the response from 604 people. These responses are developed into a LRFM model to examine customer acquisition behaviour using K-Means algorithm.

Step 2: Carrying out of LRFM analysis

Length: It is the quantity of days since the session of the purchase. The period is from the month March to August in which corona pandemic arises.

Recency: It is the quantity of time prior to the date which is referred at what time a client done their acquire products recently. If the value of recency is low, it shows that the customer to a store is high. In this graph, the data plotted are based on the recent purchase of the customers during the lockdown period.
Figure 1: Recent purchase of customers

Frequency: It refers to the time among 2 consequent purchases of a customer. The customer visit to the company high, when the value of frequency is high. In this graph, the data plotted are based on the frequent purchase of the customers during the lockdown period.

Figure 2: Frequent purchase of customers

Monetary: It is defined as the quantity of funds used up with a buyer at an exact era of occasion. If the worth of purchase is high it indicates that the company is benefited by that specific customer. In this graph, the data plotted are based on the purchase amount of the products during the lockdown period.

Figure 3: Amount spent on online shopping
Step 3:
Here for the analysis k-means Clustering algorithm is implemented. It involves in dividing the set of data into pre-planned dissimilar not intersected subsets known as Clusters in which every data point included to a single specific category. Each cluster is given a statistic point in a method that summation of the squared distance among the statistical points as well as centroid of the cluster is low. The Steps for the calculation of algorithm is listed below

The purpose of the Function is

\[ J = \sum_{i=1}^{m} \sum_{k=1}^{k} w_{ik} \| x_i - \mu_k \|^2 \]  \hspace{1cm} (1)

The E-step Comprises of

\[ \frac{\partial J}{\partial w_{ik}} = \sum_{i=1}^{m} \sum_{k=1}^{k} w_{ik} \| x_i - \mu_k \|^2 \]

\[ \Rightarrow w_{ik} = \begin{cases} 1, & \text{if } \argmin_j \| x_i - \mu_j \|^2 \\ 0, & \text{Otherwise} \end{cases} \]  \hspace{1cm} (2)

The M step is achieved through

\[ \frac{\partial J}{\partial w_{ik}} = 2 \sum_{i=1}^{m} w_{ik} (x_i - \mu_k) = 0 \]

\[ \Rightarrow \mu_k = \frac{\sum_{i=1}^{m} w_{ik} x_i}{\sum_{i=1}^{m} w_{ik}} \]  \hspace{1cm} (3)

The Final step is

\[ \frac{1}{m_k} \sum_{d=1}^{m} \| x_i - \mu_k \|^2 \]  \hspace{1cm} (4)

In the K-means Clustering algorithm we use a method named silhouette analysis to decide the degree of partition between clusters.

The steps involved are:
The typical distance from all statistical points in the same cluster is calculated as A and closest cluster is calculated as B.
With the data cooperative should be calculated by using formulae.

\[ b - a / \text{Max}(a, b) \]

The coefficient will lie between -1 to 1.
If the Coefficient is,
• 0, then model is very near to the adjacent clusters.
• 1, then model is distant from the adjacent clusters.
• -1, then model is in the incorrect clusters.
Table 1: Table Indicating LRFM value

| Index | Customer ID | Length | Recency | Frequency | Monetary |
|-------|-------------|--------|---------|-----------|----------|
| 1     | 12543       | 90     | 10.3    | 52        | 785.5    |
| 2     | 14321       | 90     | 2.6     | 39        | 1023.6   |
| 3     | 11254       | 90     | 123.5   | 63        | 1787.5   |
| 4     | 17653       | 90     | 87.9    | 110       | 174.87   |

From Table 1, we can get the data of the length, recency, frequency and monetary for each user who are using the website .for each customer an id is provided and their LRFM values are calculated based on their recent purchase, frequency of purchasing and the amount they spent to buy the products.

Silhouette Analysis

Silhouette Score for k =2 and k=4

Table 2: Silhouette Score

| Clusters | Length_log | Recency_log | Frequency_log | Monetary_log |
|----------|------------|-------------|---------------|--------------|
| 0        | 214.34     | 928.190     | 168.90        | 290.87       |
| 1        | 311.98     | 891.87      | 235.67        | 180.90       |
| 2        | 514.90     | 90.232      | 397.28        | 420.56       |

Silhouette score ,When k=2 0.387421

| Clusters | Length_log | Recency_log | Frequency_log | Monetary_log |
|----------|------------|-------------|---------------|--------------|
| 0        | 108.9      | 255.02      | 7.8719        | 180.278      |
| 1        | 206.3      | 21.367      | 92.176        | 1964.2137    |
| 2        | 361.3      | 13.456      | 87.687        | 2020.2367    |

Silhouette score ,When k=4 0.32195

From Table 2, for the Clusters 0, 1 and 2, the log values of length, recency, frequency, monetary are calculated with determining the k value as 2 and k value as 4. From the values calculated we can observe that the result we got when k=2 is more than the value obtained when k=4. That is Silhouette Score is less when the value of k is high and high when the value of k is low.
In figure 4 and 5, the image of Silhouette scheme for k=2 and k=4 be drawn respectively. The silhouette diagram comprises of the cluster labels, silhouette coefficient values, and silhouette plot for various clusters, visualization of clustered data and their eruption times in minutes the eruption data is more when the k value is more.

Step 4:
We processed this article by 3 steps. In step 1, we surveyed the reference journals then we framed the questionnaire and collected the statistics from 604 people about the online shopping. In step 2, we developed a LRFM model by relating the questionnaire with the model and then plotted the graphs using the statistics collected through the questionnaire. In step 3, we adapted the K- means cluster algorithm by which we calculated the recency and silhouette and finally plotted the graph from the gathered data.
3.2 Advantages:
The main advantages of LRFM model is to
- Increase customer retention towards the shopping
- Increase response of the customers based on their nature of buying products
- Conversion rate can be increased
- Revenue can also be extended

4. Result and discussion
The silhouette value is the computation of how alike an entity is to its own cluster compared to other cluster. Based on the range of silhouette value result is obtained in the proposed methodology. A survey is made forming a set of questionnaires in the Google forms and circulated to the users who buy online goods taking the data provided by them recency, frequency and monetary value is calculated. With the calculated value the behaviour of the customer is analysed.

5. Conclusion
One of the most accepted clustering algorithms is K-means clustering algorithm is and typically the first one the user apply when solving clustering tasks to get an thought of the arrangement of the data. In the above article, we have used the dataset collected from people about online shopping. From the developed K-mean clustering algorithm, simulated data, and the survey taken from the people through the questionnaire, it is clear that the rate of increase of online shopping is greatly increased even in the pandemic lockdown situations. Thus the success value of online shopping seems to be extended even after the pandemic. The people began to adapt to the Online shopping field than physical Shopping after this lockdown.
Since the Covid-19 Pandemic situation leads most of the people towards online shopping, the enhancement of the future the proposed methodology can be used with some of the statistical methods to bring more results with determinantal value and more accuracy. This can be even implemented to a lot of specific application such as e-commerce websites and they can use the analysis to predict the customer behaviour and value. This also helps the retail shop owners to provide best service for their best customers to increase their business and customers.

References
[1] P Anitha , Malini M and Patil2019 RFM model for customer purchase behaviour using K-Means algorithm. *Journal of King Saud University – Computer and Information Sciences*, DOI: 10.1016/j.jksuci.2019.12.011
[2] Anam Bhatti, Hamza Akram, Hafiz Muhammad Basit, Ahmed Usman Khan, SyedaMahwishRazaNaqvi, and Muhammad Bilal2020 E-commerce trends during COVID-19 Pandemic, *International Journal of Future Generation Communication*, **Vol. 13**, No. 2, PP: 1449-1452.

[3] Muhammad Kashif, Aziz-Ur-Rehman and Muhammad KashanJaved 2020Covid-19 impact on online shopping, *International Journal of Medical Science in Clinical Research and Review, Vol 3*, PP: 325-330.

[4] FeminaBahari and MSudeepElayidom 2015An Efficient CRM-Data Mining Frame work For the prediction of customer behaviour, *International Conference on Information and Communication Technologies*, DOI: 10.1016/j.procs.2015.02.136, PP: 725 – 731.

[5] DeepakArunachalam and Nirajkumar2018 Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making, *Journal of expert system application*, DOI: 10.1016/j.eswa.2018.03.007, PP: 11-34.

[6] Nicola M., Alsafi Z, Sohrab C Kerwan A, Al-Jabir A and Iosifidis C2020 The socio-economic implications of the Coronavirus and COVID-19 pandemic: A Review, *International Journal of Surgery*, DOI: 10.1016/j.ijsu.2020.04.018, PP: 185-193.

[7] Karthikeyan P A Study on Purchasing Pattern and Behaviour towards Jewellery in Erode City Managers *journal on Management,Vol.2*, Issue-10

[8] Selvaraju K and Dr.Karthikeyan P 2016 Impact on E-Commerce towards Online Shopping and Customer Buying Behavior, *Asian Journal of Research in Social*, **Vol. 6**, No.7, PP: 1260-1270.

[9] Hasanat M W, Hoque A, and Tat H H 2020, The Impact of Coronavirus on E-Business in Malaysia, *Asian Journal of Multidisciplinary Studies, Vol. 3*.

[10] Rehman S 2018 Impact of financial risk, privacy risk, convenience, and trust on online shopping with mediating role of consumer purchase intention in Pakistan, *International Journal of Academic Multidisciplinary Research, Vol. 2* Issue 8, August – 2018, PP: 27-34.

[11] Bhatti A, Saad S, and Gbadebo S M 2018 Convenience Risk, Product Risk, and Perceived Risk Influence on Online Shopping: Moderating Effect of Attitude. Science Arena Publications, *International journal of Business Management, Vol 3*, PP: 1-11.