LRFS: Online Shopper's Behaviour-Based Efficient Customer Segmentation Model

Jahnavi Settipalli
Dept of Artificial Intelligence and Data Science
Dhanalakshmi Srinivasan University, Trichy, India

Abstract- *Online shopping has become a dominant platform in modern digital commerce. Understanding customer behavior is essential for improving marketing strategies and increasing revenue. This paper proposes an enhanced customer segmentation model named LRFS, which extends the traditional LRF framework by introducing a new component called Staying Rate. The Staying Rate represents the relationship between user engagement and revenue generation. The model utilizes unsupervised machine learning techniques including K-Means and K-Medoids clustering along with dimensionality reduction methods such as PCA, t-SNE, and Autoencoder. Comparative analysis is performed between LR, LF, LRF, and the proposed LRFS model. Experimental results show that LRFS provides more accurate and meaningful customer segmentation, helping businesses identify high-value customers and improve decision-making strategies.*

Keywords: Customer segmentation, Unsupervised learning, K-Means, K-Medoids, PCA, t-SNE, Autoencoder, Google Analytics

I. INTRODUCTION

E-commerce has significantly transformed global purchasing behaviour. Businesses rely heavily on data analytics to understand customer patterns and improve engagement strategies. Customer segmentation plays a vital role in identifying groups of customers with similar behaviour.

Traditional models such as RFM analyze Recency, Frequency, and Monetary value. However, these models do not fully capture user engagement behaviour in online platforms. To address this limitation, this study proposes an improved LRFS model, which integrates Length, Recency, Frequency, and a newly introduced Staying Rate component.

The main objective of this work is to enhance segmentation accuracy using behavioural data extracted from e-commerce sessions.

1.1 Problem Statement

Although customer segmentation techniques have been widely used, traditional models have several limitations. Many existing segmentation methods focus primarily on purchase history and frequency of transactions while ignoring user engagement behaviour on websites.

In online shopping platforms, customers may spend significant time exploring products without immediately completing a purchase. Such engagement patterns provide valuable insights into customer interests and potential buying behaviour.

Therefore, there is a need for an improved segmentation model that considers both customer interaction behaviour and revenue contribution**.** The LRFS model addresses this limitation by introducing the Staying Rate component, which measures customer engagement based on website interaction data.

1.2 Project Objective

The main objective of this research is to develop a machine learning-based customer segmentation framework that improves the identification of valuable customers in e-commerce platforms.

The specific objectives include:

A. Development of an Enhanced Customer Segmentation Model

The primary objective of this project is to design and implement the **LRFS customer segmentation model**, which extends the traditional **LRF (Length, Recency, Frequency)** framework by introducing an additional component called **Staying Rate (S)**. This new feature captures customer engagement by measuring how long users remain active on the website and how this engagement contributes to revenue generation. By integrating engagement-based metrics with traditional behavioural indicators, the LRFS model aims to provide more accurate and meaningful segmentation results

B. Analysis of Online Shopper Behaviour

Another key objective of this study is to analyze the behavioural patterns of online shoppers using website interaction data. Online shopping platforms collect large volumes of customer activity data such as page visits, browsing duration, product views, and transaction records. This project aims to examine how these behavioural factors influence purchasing decisions and how they can be used to classify customers into different segments based on their activity levels and engagement patterns.

C. Application of Machine Learning Techniques for Segmentation

The project aims to apply unsupervised machine learning algorithms to automatically group customers based on their behavioural characteristics. Clustering algorithms such as K-Means and K-Medoids are used to identify hidden patterns in customer data and create meaningful clusters. These algorithms help classify customers into categories such as highly engaged customers, potential buyers, and low-engagement users, which can provide valuable insights for business decision-making.

D. Integration of Dimensionality Reduction Techniques

Customer datasets often contain high-dimensional features that may affect clustering performance and computational efficiency. Therefore, this research aims to apply dimensionality reduction techniques such as Principal Component Analysis (PCA), t-Distributed Stochastic Neighbour Embedding (t-SNE), and Autoencoder models to reduce data complexity while preserving important information. These techniques help improve cluster visualization and enhance the accuracy of segmentation results

II. LITERATURE SURVEY

Customer segmentation has become an essential technique in modern marketing and data analytics. With the rapid growth of e-commerce platforms, businesses collect large volumes of customer interaction data that can be analyzed to understand purchasing behavior and improve marketing strategies. Researchers have proposed various models and machine learning techniques to analyze customer behavior and classify users into meaningful groups.

2.1 Customer Segmentation in E-Commerce

Customer segmentation refers to the process of dividing customers into groups based on shared characteristics such as purchasing patterns, browsing behaviour, and engagement levels. In traditional marketing systems, segmentation was primarily based on demographic and geographic attributes. However, with the development of data analytics and machine learning technologies, behavioural segmentation has become more effective in identifying customer preferences.

E-commerce platforms generate large datasets containing information about user sessions, page views, product interactions, and transaction history. Analyzing these datasets helps businesses understand customer interests and predict purchasing intentions. Customer segmentation techniques allow companies to identify loyal customers, frequent buyers, and potential customers who may contribute to future revenue growth.

2.2 Traditional RFM and LRF Models

One of the most widely used models in customer segmentation is the RFM (Recency, Frequency, Monetary) model. This model evaluates customers based on three important behavioural metrics: how recently a customer made a purchase, how frequently the customer interacts with the platform, and the amount of money spent on purchases. RFM analysis helps businesses identify high-value customers and design targeted marketing campaigns.

Although RFM analysis is widely used, it has certain limitations. It focuses mainly on transaction-based metrics and may not capture the complete behavioural patterns of customers, especially in online shopping environments where browsing behaviour also plays a crucial role.

To address this limitation, researchers introduced an extension called the LRF model, where Length (L) represents the duration of the relationship between the customer and the platform. This feature helps distinguish between short-term visitors and long-term customers. By combining Length with Recency and Frequency, the LRF model provides better insights

2.3 Machine Learning Approaches for Customer Segmentation

With the advancement of machine learning techniques, customer segmentation has evolved beyond traditional statistical models. Unsupervised learning algorithms are widely used for discovering patterns in customer data without predefined labels.

Among these algorithms, K-Means clustering is one of the most commonly used techniques for customer segmentation. K-Means divides customers into clusters based on similarity in feature values, allowing businesses to identify different behavioural groups. The algorithm works by minimizing the distance between data points and cluster centroids.

Another clustering method used in segmentation studies is K-Medoids clustering, which is similar to K-Means but uses actual data points as cluster centres instead of calculated centroids. This approach is often more robust to noise and outliers in datasets.

Researchers have also explored advanced clustering techniques such as hierarchical clustering, density-based clustering (DBSCAN), and fuzzy clustering methods to improve segmentation accuracy and identify complex patterns in customer behaviour

2.4 Research Gap and Need for LRFS Model

Although several segmentation models and machine learning techniques have been proposed, many existing approaches focus mainly on purchase history and transaction frequency. These models often ignore customer engagement behaviour, which plays a significant role in online shopping environments.

In e-commerce platforms, customers frequently browse products, compare prices, and explore categories before making purchasing decisions. Metrics such as page views, session duration, and exit rates provide valuable information about user engagement and purchase intention.

III. PROPOSED METHODOLOGY

The proposed LRFS framework is designed to analyze online shopper behaviour and perform customer segmentation using machine learning techniques. The system processes customer interaction data from e-commerce platforms and groups customers based on behavioural patterns. The methodology integrates data preprocessing, feature extraction, dimensionality reduction, and clustering techniques to produce accurate and meaningful customer segments.

The overall objective of the proposed system is to identify customer groups that share similar engagement patterns and purchasing behaviours. These insights help businesses improve marketing strategies, customer retention, and personalized recommendation systems.

A. System Architecture and Workflow

The LRFS customer segmentation system follows a structured workflow consisting of multiple stages. Each stage is designed to ensure accurate data processing and effective clustering results.

The major stages of the system include:

1. **Data Collection**
2. **Data Preprocessing**
3. **Feature Engineering and LRFS Calculation**
4. **Feature Scaling**
5. **Dimensionality Reduction**
6. **Clustering and Customer Segmentation**
7. **Cluster Evaluation and Analysis**

The workflow begins with collecting customer interaction data from e-commerce platforms. The data is then cleaned and transformed before applying machine learning algorithms. After feature extraction and dimensionality reduction, clustering algorithms are used to identify meaningful customer segments.

B. Data Collection

The first step in the methodology is collecting customer behavior data from online shopping platforms. This dataset contains various attributes that describe how users interact with the website during a shopping session.

Typical features in the dataset include:

- Number of page visits
- Exit rate
- Page value
- Visitor type
- Session duration
- Month of visit
- Revenue generation

These features represent customer browsing activity and purchasing behaviour. The collected dataset serves as the foundation for performing segmentation analysis

C. Data Preprocessing

Raw datasets often contain missing values, inconsistent records, and irrelevant attributes. Therefore, data preprocessing is required to prepare the dataset for machine learning algorithms.

The preprocessing stage includes the following steps:

Data Cleaning:

Duplicate records and inconsistent values are removed to ensure data quality.

Handling Missing Values:

Missing values in numerical attributes are replaced using statistical methods such as mean or median values.

Encoding Categorical Variables :

Categorical attributes such as visitor type and month are converted into numerical form using encoding techniques such as label encoding or one-hot encoding.

Feature Scaling :

Since clustering algorithms are sensitive to the scale of data, numerical features are normalized using standard scaling techniques.

These preprocessing steps help improve clustering accuracy and reduce noise in the dataset.

D. LRFS Feature Engineering

The LRFS model introduces a behavioural segmentation framework based on four key metrics:

Length (L)

Length represents the duration of the relationship between the customer and the online platform. It indicates how long a customer has been interacting with the website over time. Customers with longer interaction periods are often considered more loyal users.

Recency (R)

Recency measures how recently a customer visited the platform. Customers who visited recently are more likely to make purchases compared to those who have not interacted with the platform for a long time.

Frequency (F)

Frequency indicates the number of times a customer interacts with the website within a specific period. Higher frequency suggests stronger engagement and increased probability of purchase.

2)Staying Rate (S)

The Staying Rate is the key contribution of the LRFS model. It measures the relationship between customer engagement and revenue generation. Customers who spend more time exploring products and pages tend to have a higher probability of making purchases.

The Staying Rate is calculated using the formula:

$$\mathbf{S = Page\ Value \times (1 - Exit\ Rate)}$$

Where:

Page Value represents the average revenue generated from a page
Exit Rate represents the percentage of users who leave the website from that page

This metric captures customer engagement and helps identify users who show strong purchase intent.

E. Dimensionality Reduction

Customer datasets often contain multiple features that increase data complexity. Dimensionality reduction techniques are applied to transform high-dimensional data into a lower-dimensional space while preserving important information.

The proposed system applies the following dimensionality reduction techniques:

Principal Component Analysis (PCA)

PCA transforms the original dataset into a set of orthogonal components that capture the maximum variance in the data. This helps reduce data dimensionality while preserving key information.

t-Distributed Stochastic Neighbor Embedding (t-SNE)

t-SNE is used for visualizing high-dimensional data by mapping it into a two-dimensional space. It helps identify cluster structures and relationships between customer groups.

Autoencoder

Autoencoders are neural network-based models used to learn compact representations of data. They reduce data dimensionality while maintaining important feature relationships.

These techniques improve clustering performance and help visualize customer segments.

F. Clustering and Customer Segmentation

After feature extraction and dimensionality reduction, clustering algorithms are applied to group customers based on their behavioural characteristics.

Two clustering algorithms are used in this study:

K-Means Clustering :

K-Means clustering partitions the dataset into K clusters by minimizing the distance between data points and cluster centroids. The algorithm iteratively updates cluster centres until convergence is achieved.

K-Medoids Clustering :

K-Medoids clustering is similar to K-Means but uses actual data points as cluster centres instead of centroids. This approach is more robust to noise and outliers in the dataset.

The optimal number of clusters is determined using techniques such as:

- Elbow Method
- Silhouette Score

These methods help identify the best cluster configuration for customer segmentation.

G. Cluster Evaluation and Analysis

Once clustering is completed, the resulting clusters are analyzed to understand customer behaviour patterns.

Each cluster represents a specific group of customers with similar characteristics. These groups may include:

- High-value customers
- Frequent buyers
- Potential buyers
- Low-engagement users

Cluster analysis helps businesses understand customer preferences and develop targeted marketing strategies.

Visualization techniques such as scatter plots and cluster distribution graphs are used to interpret segmentation results.

H. Advantages of the Proposed LRFS Framework

The proposed LRFS segmentation framework offers several advantages:

- Integrates engagement metrics with traditional behavioural features
- Provides more accurate customer segmentation
- Improves identification of high-value customers
- Supports data-driven marketing strategies
- Enhances business decision-making in e commerce platforms

By combining machine learning algorithms with behavioural analytics, the LRFS model provides a comprehensive solution for analyzing online shopper behaviour.

IV. MATHEMATICAL AND OPTIMIZATION FORMULATION

The LRFS customer segmentation model relies on mathematical representations of customer behavioural features and optimization techniques used in clustering algorithms. The purpose of this formulation is to ensure that customers with similar behavioural patterns are grouped together while maximizing the separation between different customer segments.

The proposed LRFS model integrates four behavioural metrics: **Length (L), Recency (R), Frequency (F), and Staying Rate (S)**. These metrics form the feature vector used for segmentation analysis.

A. **Mathematical Representation of LRFS Features**

Let the dataset contain **N customers**, where each customer is represented by a feature vector.

$$X_i = (L_i, R_i, F_i, S_i)$$

where:

L_i = Length of relationship of customer i

R_i = Recency of customer visit

F_i = Frequency of interactions
S_i = Staying rate representing engagement level

The complete dataset can therefore be represented as:

$X = \{X_1, X_2, X_3, \ldots, X_N\}$

Each customer vector represents behavioural characteristics that are used to perform clustering.

B. **Calculation of LRFS Components**

Length (L)

Length represents the duration of the relationship between a customer and the online platform.

$L_i = T_{last} - T_{first}$

Where

T_{first} = first recorded interaction of the customer
T_{last} = most recent interaction

A higher value of L indicates a long-term customer relationship with the platform.

Recency (R)

Recency measures how recently a customer visited the website.

$R_i = T_{current} - T_{last}$

where

$T_{current}$ = current time
T_{last} = last visit of the customer

Lower recency values indicate more recent customer activity.

Frequency (F)

Frequency represents the total number of interactions or visits made by the customer.

$$F_i = \sum_{k=1}^{n} V_{ik}$$

Where

V_{ik} represents the number of visits made by customer i

Staying Rate (S)

The Staying Rate measures customer engagement based on the time spent on website pages and revenue contribution.

$S_i = PV_i \times (1 - ER_i)$

where

PV_i = Page Value representing revenue generated from the page
ER_i = Exit Rate indicating the percentage of users leaving the page

This metric captures how effectively customer engagement contributes to revenue generation.

C. Feature Vector Normalization

Since LRFS features may have different scales, normalization is applied to ensure equal contribution of each feature in clustering.

The normalized feature value is calculated as:

$$X' = \frac{X - \mu}{\sigma}$$

where

μ = mean of the feature
σ = standard deviation

Normalization ensures that clustering algorithms are not biased toward features with larger numerical ranges.

D. Clustering Optimization Using K-Means

The objective of clustering is to partition the dataset into **K groups** such that customers within the same cluster have similar behavioural patterns.

Let:

$C = \{C_1, C_2, \ldots, C_k\}$

represent the set of clusters.

The optimization objective of K-Means clustering is to minimize the **within-cluster sum of squares (WCSS)**.

$$J = \sum_{j=1}^{k} \sum_{x_i \in C_j} || x_i - \mu_j ||^2$$

where

x_i = customer data point

μ_j = centroid of cluster C_j

The algorithm iteratively updates cluster centroids until the objective function is minimized.

E. Silhouette Score for Cluster Validation

To evaluate clustering quality, the **Silhouette Score** is used.

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

where

$a(i)$ = average distance between point i and other points in the same cluster

$b(i)$ = average distance between point i and points in the nearest cluster

The Silhouette Score ranges between **-1 and 1**, where higher values indicate better cluster separation.

F. Overall Optimization Objective

The final goal of the LRFS segmentation framework is to find cluster assignments that minimize intra-cluster distance and maximize inter-cluster separation.

The overall optimization objective can be expressed as:

$$\min J = \sum_{j=1}^{k} \sum_{x_i \in C_j} || x_i - \mu_j ||^2$$

Subject to:

$x_i \in C_j$

This optimization ensures that customers with similar engagement and purchasing behaviours are grouped together effectively.

G. Significance of the Mathematical Framework

The mathematical formulation of the LRFS model provides a structured approach for analyzing customer behaviour in e-commerce systems. By combining behavioural metrics with clustering optimization techniques, the framework enables accurate identification of customer segments.

This mathematical representation improves segmentation accuracy and helps businesses identify high-value customers based on both engagement behaviour and purchasing activity.

V. EXPERIMENTAL RESULTS AND ANALYSIS

This section presents the experimental evaluation of the proposed LRFS customer segmentation model. The analysis focuses on examining the behavioural characteristics of online shoppers and evaluating the effectiveness of the LRFS framework in identifying meaningful customer segments. The performance of the proposed model is analyzed using clustering techniques and visualization methods.

The experiments were conducted using Python and data analysis libraries such as Pandas, NumPy, and Scikit-learn. Visualization tools such as Matplotlib and Seaborn were used to present the results in graphical form.

A. dataset description

The dataset used for this experiment contains online shopping session data representing customer interactions with an e-commerce platform. Each record represents a user session and contains information related to browsing behaviour and purchasing activity.

Important attributes in the dataset include:

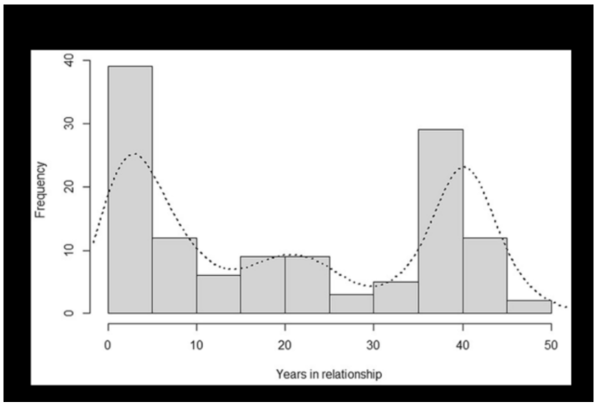
- Administrative page visits
- Informational page visits
- Product-related page visits
- Page values representing revenue contribution
- Exit rate indicating page abandonment
- Visitor type (new or returning customer)
- Month of interaction
- Revenue generated from the session

The dataset was pre-processed before analysis by removing duplicate values, handling missing data, encoding categorical variables, and normalizing numerical features.

B. Feature Distribution Analysis

Understanding the distribution of LRFS features helps analyze customer engagement patterns.

Length Distribution: Length represents the duration of the relationship between customers and the platform. Customers with longer relationships are typically more familiar with the website.



Frequency Distribution:

Frequency measures the number of visits made by customers

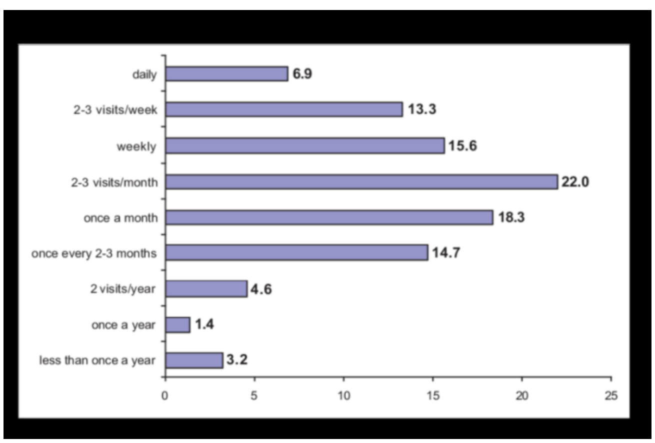


Fig: Distribution of customer visit frequency (F)

The distribution shows that a large portion of users interact occasionally, while a smaller group of highly active users frequently visit the platform.

C. Correlation Analysis of LRFS Features

A correlation heatmap was generated to analyze the relationships between Length, Recency, Frequency, Staying Rate, and Revenue.

The analysis shows that:

- Staying Rate has a strong positive relationship with revenue generation.
- Frequency also shows a moderate correlation with revenue.
- Recency demonstrates an inverse relationship with purchasing activity.

These findings indicate that engagement behavior plays an important role in determining customer purchase decisions.

D. Dimensionality Reduction Results

To improve clustering performance and visualization, dimensionality reduction techniques were applied to the LRFS dataset.

PCA Results

Principal Component Analysis was used to transform the high-dimensional dataset into two principal components. The PCA visualization revealed clear separation between customer groups, indicating the presence of distinct behavioral patterns.

t-SNE Visualization

The t-SNE algorithm was applied to visualize customer clusters in a two-dimensional space. The visualization clearly showed the formation of distinct clusters representing different customer engagement levels.

Autoencoder Representation

Autoencoder models were used to learn compact feature representations. The compressed feature space improved clustering performance by reducing noise and preserving meaningful relationships between features.

E. Clustering Results

Clustering algorithms were applied to segment customers based on LRFS features.

K-Means Clustering :

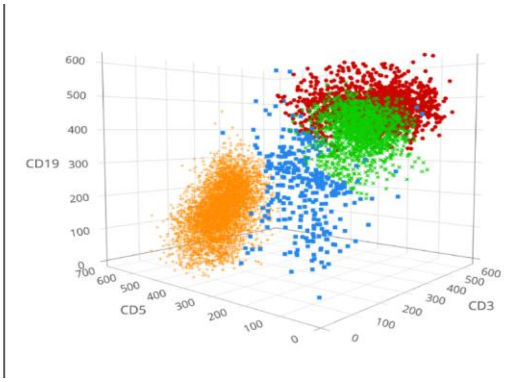


Fig: Customer segmentation using K-Means clustering.

The clustering results reveal four main customer groups:

- Highly Engaged Customers
  Customers with high frequency and high staying rate.potential Buyers
  Customers who frequently browse products but make fewer purchases.
- Occasional Visitors
  Users with moderate engagement levels.
- Inactive Customers
  Customers with very low interaction levels.

Overall Analysis

The experimental results demonstrate that the LRFS model effectively captures customer engagement behavior and improves segmentation quality.
Key findings include:

- Engagement metrics significantly influence purchasing behavior.
- Dimensionality reduction techniques improve cluster visualization.
- The LRFS model identifies high-value customer segments more effectively than traditional segmentation models

These insights help businesses develop personalized marketing strategies and improve customer retention.

VI. DISCUSSION

The experimental evaluation of the LRFS model demonstrates the effectiveness of incorporating engagement-based metrics into customer segmentation analysis. Traditional customer segmentation models often rely primarily on transaction-based attributes such as recency and frequency of purchases. While these metrics provide useful insights, they may not fully capture the behavioural patterns of users in modern e-commerce platforms where browsing activity plays a crucial role in purchasing decisions.

The introduction of the Staying Rate (S) component in the LRFS model provides a more comprehensive understanding of customer engagement. Staying Rate reflects how actively users interact with the website by considering both page value and exit rate. Customers who spend more time exploring products and pages often demonstrate a higher probability of making purchases in the future.

The clustering results obtained using K-Means and K-Medoids algorithms revealed distinct customer groups with varying levels of engagement and purchasing activity. Highly engaged customers were identified as the most valuable segment, as they exhibited high frequency of visits and strong revenue contribution. On the other hand, occasional visitors and inactive users were identified as segments with lower engagement levels.

Dimensionality reduction techniques such as PCA, t-SNE, and Autoencoder also played a significant role in improving clustering performance. These techniques helped reduce data complexity and made it easier to visualize cluster structures. The visualizations confirmed that the LRFS model produced well-separated clusters, indicating effective segmentation of customer groups.

From a business perspective, the results highlight the importance of understanding user engagement behaviour in addition to purchase history. Businesses can use the insights obtained from the LRFS model to design personalized marketing campaigns, recommend relevant products, and improve customer retention strategies. For example, highly engaged customers can be targeted with loyalty rewards, while potential buyers can be encouraged through promotional offers and personalized advertisements.

Overall, the discussion of experimental results indicates that the LRFS model provides a more effective and comprehensive approach to customer segmentation compared to traditional methods.

VII. CONCLUSION AND FUTURE SCOPE

A.CONCLUSION

This research proposed an enhanced customer segmentation framework called **LRFS (Length, Recency, Frequency, Staying Rate)** for analyzing online shopper behavior. The model extends traditional segmentation approaches by incorporating an engagement-based metric known as Staying Rate, which captures the relationship between customer interaction and revenue generation.

The proposed framework integrates multiple machine learning techniques including **K-Means clustering, K-Medoids clustering, PCA, t-SNE, and Autoencoder** to analyze customer interaction data and identify meaningful customer segments. Experimental analysis demonstrated that the LRFS model provides more accurate segmentation results compared to traditional models such as LR and LRF.

The results showed that engagement metrics significantly influence customer purchasing behavior. Customers with higher staying rates and frequent interactions were more likely to contribute to revenue generation. By identifying these high-value customer segments, businesses can develop more targeted marketing strategies and improve customer relationship management.

The LRFS framework provides a practical solution for e-commerce platforms to better understand customer behavior and support data-driven decision making. The integration of engagement-based analytics with machine learning techniques makes the model suitable for real-world applications in online retail environments.

B. FUTURE SCOPE

Although the proposed LRFS framework demonstrates promising results in customer segmentation, several improvements can be explored in future research.

One possible extension of this work is the integration of **real-time data processing systems**. Modern e-commerce platforms generate continuous streams of customer interaction data. Implementing the LRFS model in a real-time analytics environment would enable businesses to monitor customer behaviour dynamically and respond quickly to changing engagement patterns.

Another potential improvement involves the use of **deep learning-based clustering techniques**. Advanced neural network models such as deep embedding clustering and self-organizing maps could further enhance segmentation accuracy by capturing complex relationships between customer behaviour features.

Future research may also include the incorporation of **additional behavioural features**, such as clickstream data, browsing sequences, product preferences, and time spent on specific product categories. These features could provide deeper insights into customer interests and purchasing intentions.

The LRFS framework can also be extended to support **personalized recommendation systems**. By identifying customer segments with similar behavioural patterns, businesses can recommend products that match user preferences, thereby improving user experience and increasing conversion rates.Finally, the proposed model could be applied to different domains beyond e-commerce, such as **digital marketing analytics, subscription-based services, online education platforms, and financial services**, where understanding user engagement is essential for improving customer satisfaction and retention.

REFERENCES

[1] R. H. Khan, M. A. Rahman, and M. S. Islam, "LRFS: Online shoppers' behavior-based efficient customer segmentation model," *IEEE Access*, vol. 8, pp. 115812–115823, 2020.

[2] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2012.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2009.

[5] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[6] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, NY, USA: Wiley, 1990.

[7] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A*, vol. 374, no. 2065, 2016.

[8] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[9] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[10] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, NY, USA: Wiley, 1990.

[11] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A*, vol. 374, no. 2065, 2016.

[12] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
[13] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2014.
[14] P. Kotler and K. Keller, *Marketing Management*, 15th ed. Pearson Education, 2016.
[15] R. Agrawal, T. Imielinski, and A. Swami, “Mining association rules between sets of items in large databases,” in *Proc. ACM SIGMOD Int. Conf. Management of Data*, 1993, pp. 207–216.
[16] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms,” in *Proc. 10th Int. World Wide Web Conf.*, 2001, pp. 285–295.