*International Research Journal of Innovations in Engineering and Technology (IRJIET)*
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 17-24, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202503
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

# Customer-Based Market Segmentation in E-Commerce Using Hybrid Clustering

**[1]Dr. R.M.Mallika, [2]Goli Madhuri, [3]P.M.Lokesh, [4]B. Pavan Kalyan, [5]D Mahesh Babu, [6]P.Neelakumar**

[1]Hod & Professor, Department of CSE, Siddharth Institute of Engineering & Technology, Puttur, AP, India
[2,3,4,5,6]Department of CSE (AI & ML), Siddharth Institute of Engineering & Technology, Puttur, AP, India
E-mails: [1]mallika.521@gmail.com, [2]madhurigoli02@gmail.com, [3]pmlokesh2002@gmail.com, [4]bodireddypavanreddy@gmail.com, [5]maheshbabudodapaneni@gmail.com, [6]neelakumaryadav5@gmail.com

***Abstract* - Utilizing predictive modeling and data mining, this study develops market and customer segments for effective marketing strategies. Market segmentation recognizes that customers have different interests, buying habits, and preferences. By creating specialized strategies for specific target groups, a company can enhance its resource management and sales. Customer segmentation involves clustering individuals with similar characteristics and behaviors, enhancing understanding of customers' demographics and dynamic behavior.**

**The RFM (Recency, Frequency, Monetary) approach is simple and efficient for dividing markets. RFM analysis examines how recently, frequently, and financially customers make purchases, providing insights into consumer behavior. This study makes use of data mining techniques to categorize products based on recent sales, frequency of sales, and total amount spent.**

**A novel k-Means methodology for RFM analysis is introduced, aiming to improve customer segmentation and lower marketing expenses while raising customer satisfaction. The output is compared with existing RFM models, assessing the efficiency of the suggested methodology.**

**Overall, predictive modeling and BD are leveraged to create targeted marketing initiatives based on customer segmentation, ultimately enhancing sales efforts and resource allocation for companies, particularly in e-commerce platforms.**

***Keywords:*** Predictive Modelling, Customer Segmentation, RFM Analysis, Data Mining, K-Means Clustering.

## I. INTRODUCTION

Effective customer segmentation is vital for business to gain a comprehensive understanding of their customers and plan for the marketing procedures and processes accordingly. Clustering-based techniques have high demand for categorizing customers into distinct groups based on shared characteristics. These algorithms analyse customer data to identify clusters that exhibit similar attributes. Commonly used data mining mechanisms for customer segmentation include K-Means, DBSCAN, Hierarchical Clustering, and RFM Analysis. Leveraging clustering based techniques enables businesses to uncover valuable insights, develop targeted strategies, and enhance customer satisfaction and loyalty. This paper delves into the fundamental principles, advantages, and challenges of clustering algorithms, while also providing real world examples showcasing successful implementations. The subsequent sections of the paper commence with a thorough literature study on the existing and related works focusing on clustering solutions for various customer based market segmentation problem statements. The objectives of the work are been clearly stated followed by dataset description. The proposed methodology is described stage wise which further leads to a showcase to experimental results. The paper ends with relevant findings and future directions to new researchers.

Over the years, the competition amongst businesses is increased and the large historical data that is available has resulted in the widespread use of data mining techniques in extracting the meaningful and strategic information from the database of the organisation. Data mining is the process where methods are applied to extract data patterns in order to present it in the human readable format which can be used for the purpose of decision support. According to Clustering techniques consider data tuples as objects. They partition the data objects into groups or clusters, so that objects within a cluster are similar to one another and dissimilar to objects in other clusters. Customer Segmentation is the process of division of customer base into several groups called as customer segments such that each customer segment consists of customers who have similar characteristics. The segmentation is based on the similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits. The customer segmentation has the importance as it includes, the ability to modify the

 

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 17-24, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202503
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

programs of market so that it is suitable to each of the customer segment, support in business decision; identification of products associated with each customer segment and to manage the demand and supply of that product; identifying and targeting the potential customer base, and predicting customer defection, providing directions in finding the solutions.

This template, modified in MS Word 2007 and saved as a "Word 97-2003 Document" for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

## II. LITERATURE SURVEY

An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry 2023 T. Ho, S. Nguyen, H. Nguyen, N. Nguyen, D. S. Man, and T. G. Le This study enhances the RFM model with demographic data, improving targeted marketing and business decisions.

A Study on Customer Segmentation Using K-Means Clustering for Online Shoppers 2023 R. M. Metilda This research demonstrates K-Means clustering's effectiveness for customer segmentation and personalized marketing in e-commerce.

Knowing Your Customers Using Customer Segmentation 2022 Das Parichay and Vijendra Singh This chapter explores customer segmentation, blending traditional methods with modern techniques like clustering algorithms and the RFM model to enhance marketing.

Unsupervised K-Means Clustering Algorithm 2020 K. P. Sinaga and M. S. Yang This paper presents a comprehensive analysis of the unsupervised K-Means clustering algorithm, which is widely used for customer segmentation and other data classification tasks.

Customer Segmentation Using K-Means Clustering 2018 Kansal Tushar et al. This paper explores the use of K-Means clustering for customer segmentation in various business contexts.

Customer Segmentation Using RFM Model and K-Means Clustering 2021 Shirole Rahul, Laxmiputra Salokhe, and Saraswati Jadhav This research paper focuses on customer segmentation using the RFM (Recency, Frequency, Monetary) model combined with the K-Means clustering algorithm.

## III. EXISTING SYSTEM

The existing system for customer segmentation typically relies on traditional methods like demographic profiling, surveys, and manual segmentation, which can be time-consuming and often fail to capture the nuanced purchasing behaviors of customers. Many businesses use basic segmentation models, such as geographic or age-based segmentation, which may not reflect the complexity of modern consumer behavior. These systems often overlook important factors like customer recency, frequency, and monetary value, leading to less accurate insights and ineffective marketing strategies. Additionally, conventional methods may struggle with handling large datasets, resulting in slower decision-making processes and less targeted marketing. In contrast, modern techniques like RFM analysis combined with K-Means clustering offer a more data-driven approach, enabling businesses to derive more granular insights and make faster, more informed decisions. However, the adoption of these advanced data mining techniques remains limited in many sectors due to the complexity of implementation and the need for specialized knowledge. Existing systems often lack the flexibility and scalability required to continuously adapt to shifting market dynamics, making it challenging to maintain competitive advantage in a rapidly evolving business landscape.

## IV. PROPOSED SYSTEM

The proposed system aims to enhance customer segmentation and marketing strategies for e-commerce platforms by leveraging the RFM (Recency, Frequency, Monetary) model and the k-Means clustering algorithm. It processes customer transaction data to calculate RFM metrics, categorizing customers based on their purchasing behaviors. The k-Means algorithm clusters customers into distinct segments, allowing for personalized marketing strategies tailored to different customer groups. The system, built using Java and Spring Boot, enables real-time data processing and integration with e-commerce platforms. Predictive modeling techniques are incorporated to forecast future customer behaviors, further optimizing marketing efforts. Visualization dashboards provide actionable insights into segment performance, helping businesses allocate resources effectively.

 

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 17-24, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202503
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

The system is designed for scalability and efficiency, offering significant improvements in customer targeting, sales conversion, and overall marketing ROI. Furthermore, it enables businesses to reduce marketing costs by focusing on high-value customers while enhancing customer satisfaction through personalized campaigns. By continuously analyzing and updating customer segments, the system ensures adaptive marketing strategies that keep pace with evolving consumer behaviors. This dynamic approach strengthens customer loyalty and improves overall sales performance. The system's flexibility allows it to integrate seamlessly into existing e-commerce workflows, making it a valuable tool for businesses looking to enhance their customer engagement.

## V. ARCHITECTURE

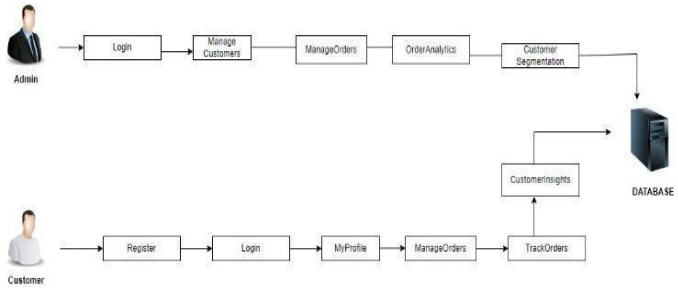


**Figure 1: Architecture of Customer-Based Market Segmentation in E-Commerce**

## VI. ALGORITHMS

### K-Means Clustering

We are using the K-Means clustering algorithm for customer segmentation in e-commerce. First, we analyze customer purchasing behavior using the RFM (Recency, Frequency, and Monetary) model, where we calculate how recently a customer made a purchase, how often they buy, and how much they spend. These three values form a dataset, which we then use as input for the K-Means algorithm. The algorithm groups customers into different clusters based on their similarities in shopping behavior. For example, it helps us identify loyal customers, frequent buyers, occasional shoppers, and inactive users. This segmentation allows businesses to implement targeted marketing strategies, personalized offers, and better resource allocation. By applying K-Means, we improve accuracy compared to traditional segmentation methods, reduce marketing costs, and increase customer engagement. Additionally, we compare the results with existing RFM models to ensure the effectiveness of our approach. Overall, this method helps businesses understand their customers better and make data-driven decisions to improve sales and retention.

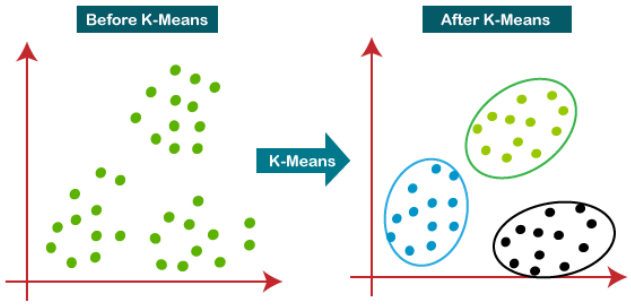


**Figure 2: K-Means Clustering**

### RFM Analysis (Recency, Frequency, Monetary)

RFM Analysis is a customer segmentation technique that evaluates customer behavior using three key metrics: Recency (how recently a customer made a purchase), Frequency (how often they make purchases), and Monetary (how much they spend). Customers are scored on each of these metrics, typically on a scale of 1 to 5, and the scores are combined to create an RFM score. This score is used to segment customers into groups, such as high-value, loyal, or at-risk customers. These segments are then used to tailor marketing strategies and improve customer engagement. RFM Analysis is widely applied in customer segmentation, targeted marketing, and customer lifetime value analysis. Its simplicity and ease of implementation make it a popular choice for businesses. However, it relies heavily on transactional data and may not capture broader customer behavior or preferences, which can limit its effectiveness in some scenarios.

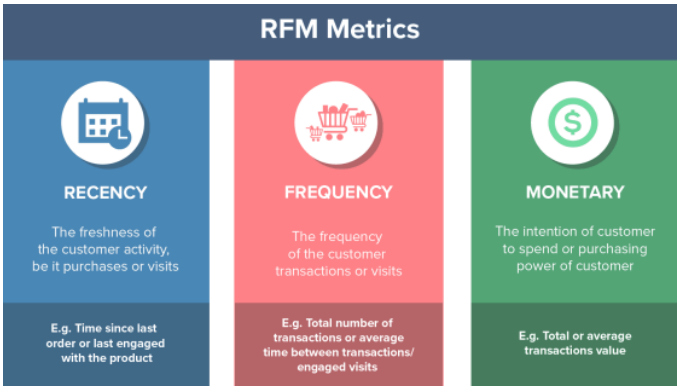


**Figure 5: RFM Analysis**

### Working

1. Data Preprocessing:

The first step involves preparing the dataset by extracting Recency, Frequency, and Monetary (RFM) values from customer transaction records. The data is cleaned by handling missing values, normalizing features, and ensuring

 

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 17-24, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202503
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

consistency. This ensures that the clustering algorithm receives structured input for effective segmentation.

2. RFM Score Calculation:

Each customer is assigned an RFM score based on:

- Recency (R): How recently a customer made a purchase.
- Frequency (F): How often a customer purchases.
- Monetary (M): How much a customer spends.

These values serve as key indicators for clustering customers based on their purchasing behavior.

3. Applying K-Means Clustering:

- The K-Means algorithm is applied to segment customers into different groups based on their RFM scores.
- Customers with similar shopping behaviors are grouped into clusters, such as loyal customers, frequent buyers, and inactive users.
- The optimal number of clusters is determined using the Elbow Method, ensuring meaningful segmentation.

4. Cluster Analysis & Insights Generation:

- The segmented data is analyzed to understand different customer groups and their purchasing habits.
- Businesses can create targeted marketing strategies tailored to each customer segment.
- High-value customers may receive exclusive offers and loyalty rewards, while inactive customers can be re-engaged through personalized promotions.

5. Model Evaluation & Performance Analysis:

- The effectiveness of K-Means clustering is assessed by using metrics such as Silhouette Score and Davies-Bouldin Index.
- The clustering results are compared with traditional segmentation methods to determine improvements in customer insights and marketing effectiveness.

6. Integration & Deployment:

- The clustering model is integrated into an admin dashboard, where businesses can visualize customer segments and insights in real-time.
- The system continuously updates clusters based on new data, enabling businesses to adapt marketing strategies dynamically as customer behaviors change.

**Objective:**

The goal of this system is to enhance customer segmentation in e-commerce by leveraging K-Means clustering on RFM data to improve personalized marketing, customer engagement, and sales performance. This approach enables businesses to optimize resource allocation and make data-driven decisions for better customer retention and revenue growth.

## VII. MODULES

**Modules:**

The proposed customer-based market segmentation system is divided into several key modules, each responsible for different functionalities. These modules ensure efficient data processing, customer segmentation, and order management. The main modules and their sub-modules are:

### A. Admin Module

The Admin Module is responsible for overseeing the entire system's operations, including customer registration, order management, and customer segmentation. It consists of the following sub-modules:

1. Customer Management Module

- Manages customer registrations, profile updates, and authentication.
- Stores and secures customer data for segmentation analysis.
- Enables admin users to view customer details and purchase history.

2. Order Processing Module

- Tracks customer orders from placement to completion.
- Manages order statuses such as pending, shipped, and delivered.
- Provides reports on order trends and customer purchasing behavior.

3. Customer Segmentation Module

- Uses the RFM (Recency, Frequency, Monetary) model to classify customers.
- Implements k-Means clustering to group customers based on their purchasing patterns.
- Dynamically updates customer segments as new transaction data is received.

 

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 17-24, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202503
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

4. Insights and Reporting Module

- Generates detailed reports on customer segments, sales trends, and marketing effectiveness.
- Visualizes data through dashboards, showing customer spending patterns and engagement levels.
- Assists in decision-making for targeted marketing and resource allocation.

5. Security and Data Handling Module

- Ensures that customer data is stored securely with encryption.
- Implements role-based access control (RBAC) for different administrative levels.
- Monitors system activity logs to detect potential security threats.

**B. Customer Module**

The Customer Module enables customers to interact with the system, browse products, place orders, and receive personalized recommendations. It consists of the following sub-modules:

1. Product Browsing and Purchase Module

- Allows customers to browse available products and view detailed descriptions.
- Supports adding products to a cart and completing purchases.
- Displays product recommendations based on purchase history and preferences.

2. Order Tracking Module

- Enables customers to view their order status, from processing to delivery.
- Provides estimated delivery timelines and real-time shipping updates.
- Allows customers to cancel or modify their orders within a set timeframe.

3. Personalized Recommendations Module

- Analyzes customer purchase history to provide personalized product suggestions.
- Offers discounts and promotions based on customer engagement and loyalty.
- Uses clustering insights from the admin module to enhance recommendations.

4. Customer Insights Module

- Displays customer segmentation details based on RFM scores.
- Provides insights into shopping habits, spending patterns, and frequency of purchases.
- Suggests loyalty programs and rewards for high-value customers.

5. Security and Privacy Module

- Ensures secure login and authentication for customer accounts.
- Encrypts sensitive customer data such as payment information.
- Allows customers to manage privacy settings and opt out of data tracking.

Each of these modules plays a crucial role in enhancing customer segmentation, optimizing marketing strategies, and improving the overall customer experience. By leveraging data mining techniques and predictive analytics, the system provides businesses with actionable insights to boost customer retention and drive sales.

## VIII. EXPERIMENTAL RESULTS

- This study utilized customer transaction data to conduct market segmentation. The dataset comprises various attributes related to purchasing behavior, such as recency, frequency, and monetary value. These features play a crucial role in clustering and segmenting customers effectively.
- To ensure a well-balanced model, the dataset was split into 70% for training and 30% for testing. This approach allows the model to learn from a significant portion of the data while evaluating its performance on unseen data to ensure reliable segmentation results.
- The k-Means clustering algorithm was employed to group customers based on their purchasing behavior. This method is widely recognized for its efficiency in segmenting data into meaningful clusters. The optimal number of clusters was identified using the elbow method, which helps determine the most suitable segmentation strategy.
- The performance of customer segmentation was assessed using various metrics, including inertia (which measures the compactness of clusters), silhouette score (which evaluates the separation between clusters), and cluster cohesion. These indicators helped validate the accuracy and effectiveness of the clustering process.
- The quality of segmentation was further analyzed by examining intra-cluster similarity and inter-cluster

 

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 17-24, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202503
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

distinction. The goal was to ensure that customers within the same segment exhibited similar purchasing patterns while maintaining clear separation between different clusters. The findings confirmed that the model successfully identified distinct customer groups, providing valuable insights for targeted marketing campaigns and personalized recommendations.

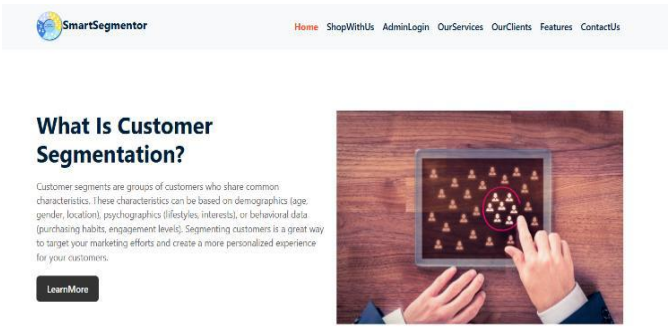


**Figure 6: Home Page**

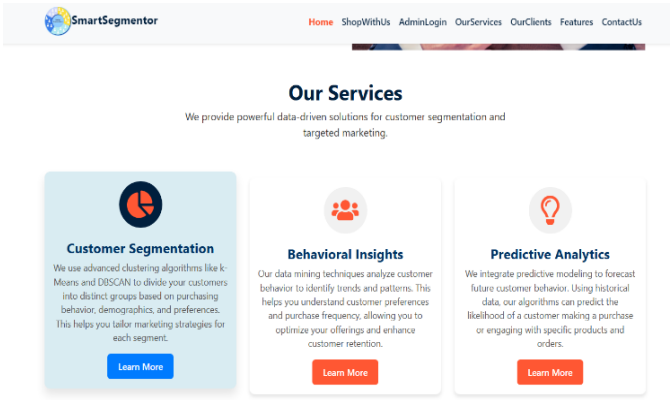


**Figure 7: Services related Information**

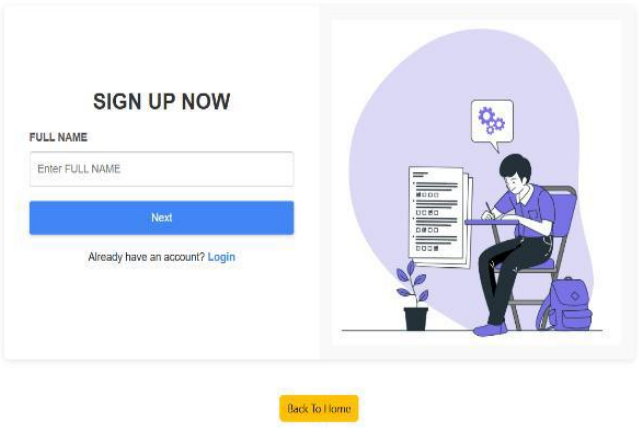


**Figure 8: Sign Up Page**

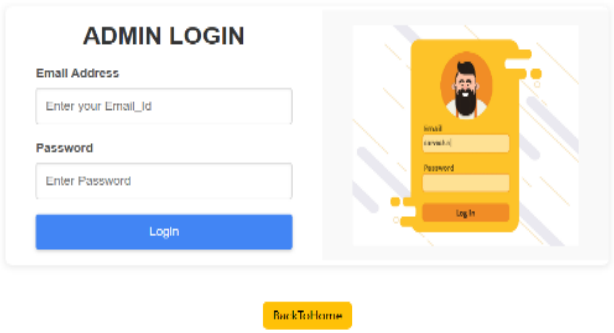


**Figure 9: Admin Login**

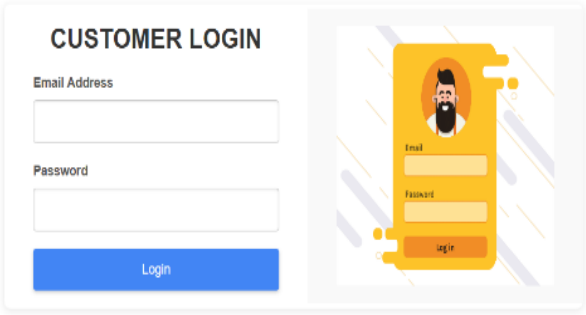


**Figure 10: Customer Login**

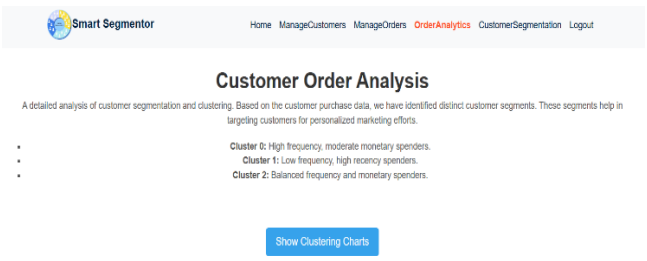


**Figure 11: Customer Order Analysis**

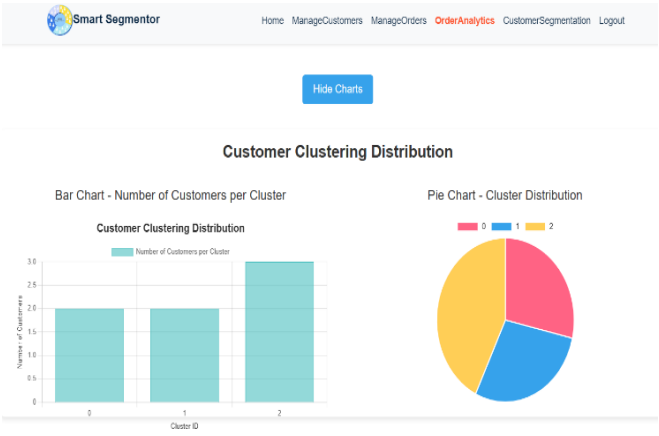


**Figure 12: Bar Chart – Customers per Cluster (30%, 20%, 50%)**

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*
**ISSN (online): 2581-3048**
**Volume 9, Special Issue ICCIS-2025, pp 17-24, May-2025**
https://doi.org/10.47001/IRJIET/2025.ICCIS-202503
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

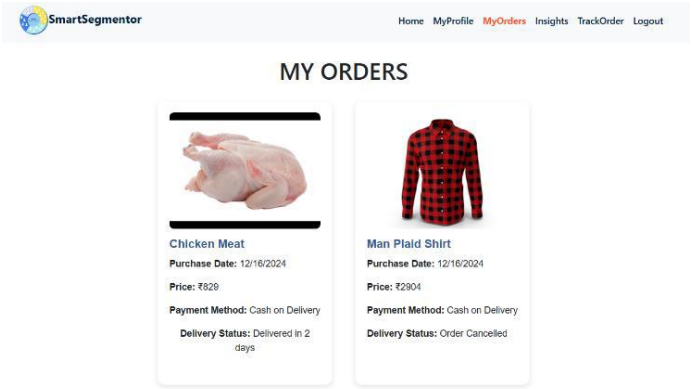


**Figure 13: Order History**

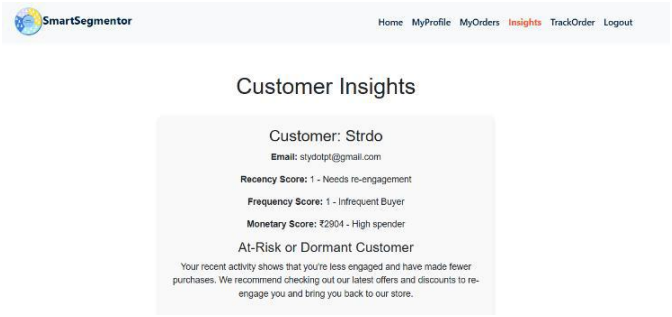


**Figure 14: Customer Insights**

## IX. CONCLUSION

The Customer-Based Market Segmentation Using Clustering in Data Mining system provides a powerful solution for businesses to understand customer behavior and optimize their marketing strategies. By leveraging the RFM analysis and k-Means clustering, the system effectively segments customers into distinct groups based on their purchasing habits, enabling businesses to create targeted marketing campaigns that enhance customer engagement and drive sales. The integration of advanced data mining techniques allows for more precise decision-making, while the Admin and Customer modules work together to improve operational efficiency and the overall shopping experience. As customer data continues to grow, future enhancements such as advanced machine learning algorithms, real-time processing, and external data integration will further refine the system, ensuring it remains adaptive and scalable. This system demonstrates the value of data-driven marketing and the potential for businesses to build stronger relationships with their customers through personalized, insights-driven strategies. By improving customer retention and optimizing resource allocation, the system ultimately helps businesses achieve higher ROI on their marketing efforts. Furthermore, the scalability of the system allows it to support businesses of varying sizes, making it a versatile tool for companies looking to improve their customer engagement and sales performance. As technology evolves, the system can adapt to emerging trends, ensuring long-term relevance and continued success in the competitive market landscape.

## X. FUTURE SCOPE

As businesses continue to accumulate more customer data, the Customer-Based Market Segmentation Using Clustering in Data Mining system can be enhanced to provide even deeper insights and more refined customer segments. One potential future enhancement is the integration of advanced machine learning algorithms for better prediction and dynamic clustering. Currently, the system uses the k-Means clustering algorithm, but the introduction of more sophisticated techniques, such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) or Hierarchical Clustering, could provide even more accurate results, especially for customers with non-linear or complex buying patterns. Additionally, incorporating predictive analytics can help forecast customer behavior, such as predicting the likelihood of a customer making a purchase or churning, allowing businesses to take preemptive actions to retain high-value customers.

Another area for enhancement is the real-time processing and integration of external data sources to provide a more comprehensive view of customer behavior. The current system primarily relies on internal transaction data, but future iterations could include social media data, web browsing behavior, and customer sentiment analysis to further refine customer segments. By integrating real-time data streams, the system could enable businesses to react quickly to changes in customer behavior and market trends, ensuring that marketing campaigns are always relevant and timely. Additionally, automation in the form of AI-driven decision-making can be implemented, where the system autonomously adjusts customer segments and marketing strategies based on the most recent data, further improving efficiency and the customer experience. These advancements would ensure the system remains adaptable, scalable, and ahead of the curve in a rapidly changing business environment.

## REFERENCES

[1] T. Ho, S. Nguyen, H. Nguyen, N. Nguyen, D. S. Man, and T. G. Le, "An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry," Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy, vol. 14, no. 1, pp. 26-53, 2023.

[2] R. M. Metilda, "A Study on Customer Segmentation Using K-Means Clustering for Online Shoppers,"

 

Rivista Italiana di Filosofia Analitica Junior, vol. 14, no. 2, pp. 1050-1058, 2023.

[3] S. Harish and C. Malathy, "Customer Segment Prediction on Retail Transactional Data Using K-Means and Markov Model," Intelligent Automation & Soft Computing, vol. 36, no. 1, 2023.

[4] J. M. John, O. Shobayo, and B. Ogunleye, "An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market," Analytics, vol. 2, no. 4, pp. 809-823, 2023.

[5] Das Parichay and Vijendra Singh, "Knowing Your Customers Using Customer Segmentation," in Computational Methods and Data Engineering: Proceedings of ICCMDE 2021, Singapore: Springer Nature Singapore, pp. 437-451, 2022.

[6] Shirole Rahul, Laxmiputra Salokhe, and Saraswati Jadhav, "Customer Segmentation Using RFM Model and K-Means Clustering," Int. J. Sci. Res. Sci. Technol, vol. 8, pp. 591-597, 2021.

[7] Li Yue et al., "Customer Segmentation Using K-Means Clustering and the Adaptive Particle Swarm Optimization Algorithm," Applied Soft Computing, vol. 113, pp. 107924, 2021.

[8] S. Koul and T. M. Philip, "Customer Segmentation Techniques on E-Commerce," in International Conference on Advanced Computing and Innovative Technologies in Engineering (ICACITE), pp. 135-138, 2021, March.

[9] K. P. Sinaga and M. S. Yang, "Unsupervised K-Means Clustering Algorithm," IEEE Access, vol. 8, pp. 80716-80727, 2020.

[10] Kansal Tushar et al., "Customer Segmentation Using K-Means Clustering," in 2018 International Conference on Computational Techniques Electronics and Mechanical Systems (CTEMS), 2018.

**Citation of this Article:**

Dr. R.M.Mallika, Goli Madhuri, P.M.Lokesh, B. Pavan Kalyan, D Mahesh Babu, & P.Neelakumar. (2025). Customer-Based Market Segmentation in E-Commerce Using Hybrid Clustering. In proceeding of Second International Conference on Computing and Intelligent Systems (ICCIS-2025), published in *IRJIET*, Volume 9, Special Issue ICCIS-2025, pp 17-24. Article DOI https://doi.org/10.47001/IRJIET/2025.ICCIS-202503

*******

 