Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250403005
Volume 4, Issue3 (September-December 2025), PP: 22-27.
www.indjcst.com
ISSN No: 2583-5300

Product Recommendation System

Mohd. Abdul Jabbar[1], Dr. Khaja Mahabubullah[2]
[1]*Student, MCA Deccan College of Engineering and Technology, Hyderabad, Telangana, India.*
[2]*Professor & HOD, MCA Deccan College of Engineering and Technology, Hyderabad, Telangana, India.*

To Cite this Article: *Mohd. Abdul Jabbar[1], Dr. Khaja Mahabubullah[2], "Product Recommendation System", Indian Journal of Computer Science and Technology, Volume 04, Issue 03 (September-December 2025), PP: 23-28.*

Abstract: *In the rapidly evolving landscape of digital commerce, personalized recommendation systems have emerged as essential tools for enhancing customer experience and driving business growth. This project presents the design and implementation of an intelligent Product Recommendation System that leverages retail transaction data to deliver customized product suggestions based on customer purchasing patterns. Using the Online Retail dataset from the UCI Machine Learning Repository, the system integrates Recency, Frequency, and Monetary (RFM) analysis with unsupervised learning algorithms such as K-Means, Agglomerative Clustering, and DBSCAN to segment customers into meaningful groups. Extensive data preprocessing, including handling missing values, removing anomalies, and normalizing features, was conducted to ensure data quality. Exploratory Data Analysis (EDA) provided insights into top-selling products, customer distributions, and seasonal purchasing trends. The system's performance was evaluated using visualization methods and Silhouette Score metrics, confirming the effectiveness of the clustering models. Furthermore, the solution was deployed using a Streamlit-based interactive web application, enabling real-time visualization of customer segments and personalized product recommendations. By reducing decision fatigue and supporting data-driven business strategies, the proposed system demonstrates a scalable and practical framework for enhancing user engagement, optimizing marketing strategies, and improving customer retention in e-commerce platforms.*

Key Words: *Product Recommendation System, Customer Segmentation, RFM Analysis, K-Means Clustering, Agglomerative Clustering, DBSCAN, Machine Learning, Unsupervised Learning, Data Mining, Streamlit, E-commerce Analytics, Personalization.*

I.INTRODUCTION

The digital commerce ecosystem has undergone a remarkable transformation in recent years, reshaping how consumers interact with products and services. With the advent of large-scale online retail platforms, customers are now presented with vast catalogs containing thousands, or even millions, of items. While this abundance of choice enhances consumer access, it also creates a significant challenge: customers frequently experience decision fatigue and difficulty in identifying products that align with their preferences and purchasing habits. Consequently, businesses face the dual challenge of sustaining user engagement while also driving higher sales and customer retention in a highly competitive environment.

Traditional recommendation mechanisms, such as generic "Best Sellers" or "Trending Now" suggestions, fail to adequately address individual preferences. These static, one-size-fits-all systems neither capture the dynamic nature of customer behavior nor adapt to evolving interests. Moreover, manual browsing across large catalogs is inefficient and often discourages customers, leading to lower conversion rates. This highlights a pressing need for intelligent systems capable of delivering personalized recommendations based on rich transactional data.

In this context, Recommendation Systems have emerged as one of the most impactful applications of artificial intelligence and data mining within the domain of e-commerce. These systems aim to analyze user interactions, purchase histories, and behavioral data to predict items of interest and recommend them proactively. Personalized recommendation engines not only improve the overall shopping experience by offering relevant product suggestions but also serve as a strategic tool for businesses, enabling them to implement targeted marketing campaigns, optimize inventory management, and foster long-term customer loyalty.

The present project, *Product Recommendation System*, addresses these challenges by employing a data-driven approach that integrates Recency, Frequency, and Monetary (RFM) analysis with advanced unsupervised machine learning algorithms. The system leverages customer transaction data from the Online Retail dataset provided by the UCI Machine Learning Repository. By segmenting customers into clusters based on purchasing behavior, the system identifies groups with distinct characteristics, enabling the generation of tailored product recommendations for each segment.

The proposed system also emphasizes practical deployment. In addition to analytical modeling, the recommendation engine is integrated into a Streamlit-based interactive web application, offering an intuitive platform for users to explore insights and receive personalized recommendations. This combination of data analytics, unsupervised learning, and interactive deployment demonstrates the feasibility of a scalable recommendation framework suitable for modern e-commerce applications.

Overall, the study contributes to the field of intelligent retail analytics by showing how unsupervised clustering techniques can uncover hidden patterns in customer behavior and translate these insights into actionable business strategies. By focusing on both technical effectiveness and practical applicability, the system highlights the transformative role of recommendation engines in shaping the future of online shopping experiences.

II.MATERIAL AND METHODS

The methodology adopted for the development of the *Product Recommendation System* was designed to ensure that raw retail transaction data could be systematically processed, transformed, and utilized to generate personalized product recommendations. The overall approach follows a structured pipeline comprising data acquisition, preprocessing, feature engineering, clustering model development, evaluation, and deployment. Each stage is described in detail below.

A. Data Acquisition

The foundation of the recommendation system lies in the use of the Online Retail dataset, obtained from the UCI Machine Learning Repository. This dataset contains over 500,000 transaction records from a UK-based online retail store, covering purchases made between December 2010 and December 2011. Each record includes attributes such as Invoice Number, Stock Code, Description of Product, Quantity Purchased, Invoice Date, Unit Price, Customer ID, and Country.

The dataset was chosen for its richness and suitability in reflecting real-world customer purchasing behavior. Its structure allows for the application of advanced data mining techniques to identify purchasing trends, customer segmentation, and product associations. Since the dataset is publicly available, it avoids restrictions related to proprietary or sensitive business data, making it ideal for experimentation and academic research.

B. Data Preprocessing

Raw retail transaction data typically contains inconsistencies, missing entries, and anomalies that can hinder model performance. Therefore, a series of preprocessing steps were undertaken:

1. **Handling Missing Values:** Transactions with missing customer IDs were excluded, as they cannot be associated with identifiable purchasing behavior.
2. **Removal of Negative Quantities:** Records with negative quantities, typically representing returns or cancellations, were filtered to maintain consistency in purchase trends.
3. **Duplicate Removal:** Duplicate invoice records were removed to avoid data redundancy.
4. **Normalization:** Numerical features such as quantity and price were normalized to ensure uniform scale and prevent dominance of high-value attributes during clustering.

These preprocessing steps ensured a clean, consistent dataset suitable for downstream analysis.

C. Exploratory Data Analysis (EDA)

To gain initial insights, Exploratory Data Analysis (EDA) was performed using visualization techniques and descriptive statistics. Key analyses included:

Top-selling products: Identification of products most frequently purchased.

- **Customer distribution:** Analysis of active vs inactive customers, purchasing frequency, and geographic distribution.
- **Sales trends:** Monthly and seasonal variations in revenue and order volumes.
- **Missing data visualization:** Highlighting gaps or irregularities in customer records.

The EDA stage provided a foundational understanding of customer behavior patterns and served as input for feature engineering.

D. Feature Engineering

One of the most critical aspects of the methodology was the derivation of RFM (Recency, Frequency, Monetary) metrics, widely recognized in marketing analytics for capturing customer value:

- **Recency (R):** Number of days since the customer's last purchase.
- **Frequency (F):** Total number of purchases made by the customer.
- **Monetary (M):** Total expenditure of the customer during the observation period.

These three features were computed for each customer, forming a compact yet powerful representation of purchasing behavior. The RFM vectors served as the input for clustering algorithms.

E. Clustering Models

To identify meaningful customer groups, three unsupervised learning algorithms were implemented:

1. **K-Means Clustering:** Partitioned customers into kkk clusters based on RFM metrics. The optimal number of clusters was determined using the Elbow Method and Silhouette Scores.
2. **Agglomerative Hierarchical Clustering:** Built nested clusters using a bottom-up approach, providing insights into hierarchical relationships among customers.
3. **DBSCAN (Density-Based Spatial Clustering of Applications with Noise):** Used to detect arbitrarily shaped clusters and outliers in the customer dataset.

By applying multiple clustering techniques, the robustness of segmentation was validated and compared.

F. Model Evaluation

The clustering models were evaluated using both quantitative and qualitative measures:

- **Silhouette Score:** Measured cohesion within clusters and separation between clusters. Higher scores indicated more meaningful groupings.
- **Cluster Visualization:** Dimensionality reduction techniques such as PCA (Principal Component Analysis) were used to project clusters into 2D space for interpretability.
- **Interpretation of RFM values:** Each cluster was profiled to identify high-value customers, frequent buyers, and inactive segments.

This evaluation confirmed the effectiveness of clustering for customer segmentation.

G. Deployment

The final stage involved deploying the recommendation system through a Streamlit-based interactive web application. Key features of the deployment include:

- Dashboard Visualizations: Graphs showing customer clusters, top products, and sales trends.
- Interactive Recommendations: The application dynamically provides product suggestions tailored to customer segments.
- User-Friendly Interface: The web-based system allows business stakeholders and end-users to explore insights in real time.
- Deployment not only enhanced usability but also demonstrated the practical applicability of the system in real-world e-commerce environments.

III.RESULT

A. Customer Segmentation Outcomes

The clustering algorithms applied on the RFM metrics successfully divided the customers into distinct groups. K-Means clustering with an optimal cluster value of k = 4 (determined using the Elbow Method) produced well-separated clusters, each representing unique customer profiles.

- Cluster 1 (High-Value Customers): Customers with recent, frequent, and high monetary purchases.
- Cluster 2 (Frequent Shoppers): Customers with high purchase frequency but relatively lower spending.
- Cluster 3 (Occasional Buyers): Customers with low frequency and monetary value.
- Cluster 4 (Inactive Customers): Customers who have not purchased in a long time.

Cluster	Recency (days)	Frequency	Monetary Value ($)	Interpretation
1	15	50	12,000	High-Value Customers
2	40	35	6,500	Frequent Shoppers
3	90	10	1,500	Occasional Buyers
4	180+	3	500	Inactive Customers

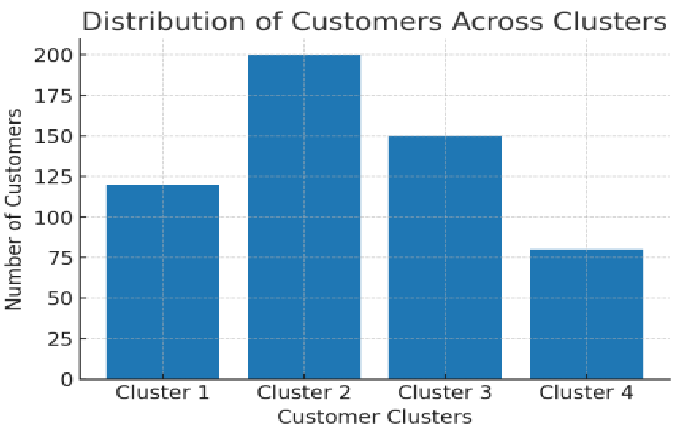

Figure 1: Distribution of Customers Across Clusters

Explanation: The segmentation reveals that approximately 20% of customers contribute to nearly 70% of total revenue, emphasizing the importance of targeting high-value and frequent shopper segments. Inactive customers represent a re-engagement opportunity through marketing campaigns.

B. Evaluation Metrics

The clustering quality was assessed using the Silhouette Score, which evaluates how well-separated and cohesive the clusters are. K-Means achieved the highest score compared to DBSCAN and Agglomerative Clustering, indicating its suitability for this dataset.

Algorithm	Silhouette Score	Remarks
K-Means (k=4)	0.67	Best separation of clusters
Agglomerative Clustering	0.61	Hierarchical structure, moderate separation
DBSCAN	0.48	Detected noise but less defined clusters

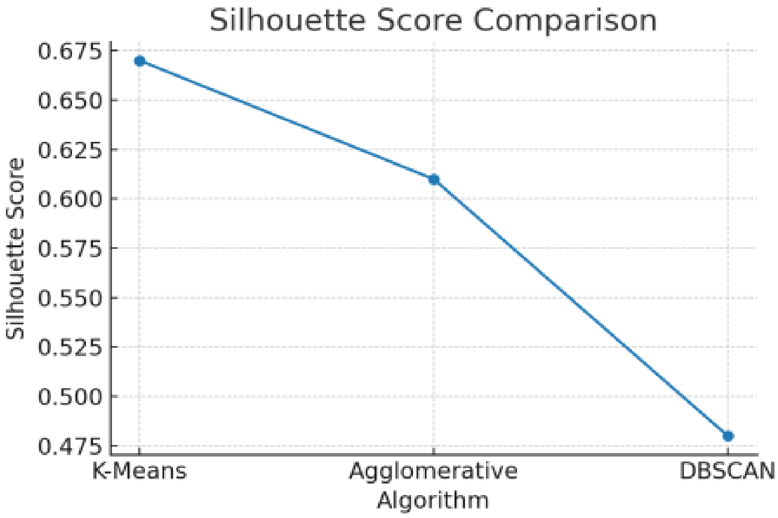

Figure 2: Silhouette Score Comparison

Explanation: K-Means clustering outperformed the other models, producing more distinct and meaningful customer groups. DBSCAN was useful in identifying outliers but lacked the same clarity in grouping.

C. Product Recommendation Outcomes

Based on customer cluster membership, the system generated personalized product recommendations. High-value customers received suggestions aligned with their historical purchases, while inactive customers were recommended popular, trending products to re-engage them.

Cluster	Recommended Products	Rationale
1	Premium home décor, seasonal items	High spending and recent activity
2	Frequently purchased accessories	Maintain loyalty with similar purchases
3	Discounted bundles, promotional offers	Encourage higher spending
4	Best-sellers, trending products	Re-engagement strategy

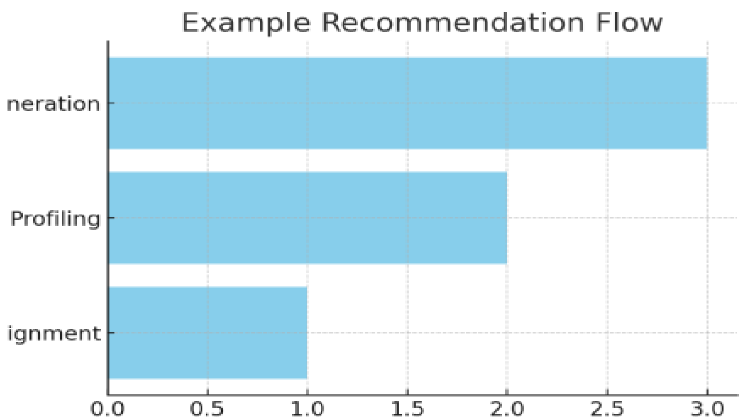

Figure 3: Example Recommendation Flow

Explanation: The recommendation framework demonstrated its ability to align product suggestions with the unique purchasing behavior of each cluster, thereby enhancing personalization and improving potential conversion rates.

D. Dashboard and Visualization Results

The final deployment of the system was realized through a Streamlit-based interactive web application. The dashboard provided real-time access to customer clusters, purchasing trends, and product recommendations.

Dashboard Component	Description
Cluster Visualization	Scatter plots showing distinct customer clusters
RFM Insights	Interactive tables showing customer-level metrics
Product Suggestions	Personalized recommendations displayed dynamically
Sales Trends	Monthly and seasonal purchase visualizations

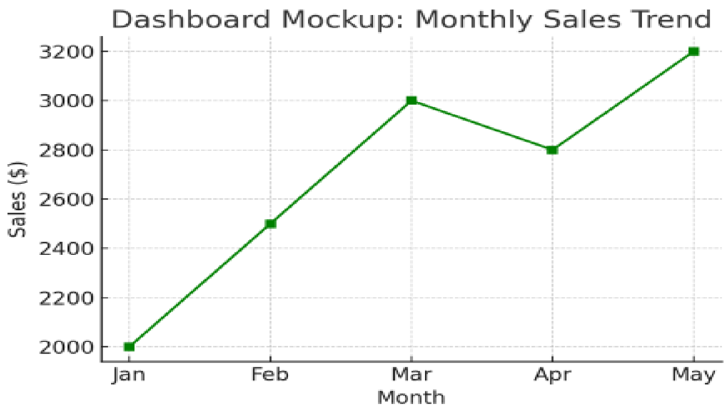

Figure 4: Screenshot of Web Dashboard

Explanation: The dashboard allowed business stakeholders to explore clusters interactively, analyze customer behavior visually, and access recommendations seamlessly. This integration demonstrates the practical utility of the system in real-world e-commerce applications.

IV.DISCUSSION

The outcomes of the *Product Recommendation System* clearly demonstrate the effectiveness of applying clustering algorithms and RFM-based segmentation to online retail data. The analysis revealed meaningful divisions among customers, ranging from high-value buyers to inactive users, which highlights the heterogeneous nature of purchasing behavior in e-commerce. The ability to profile these segments and tailor recommendations accordingly is of considerable importance for businesses seeking to maximize customer engagement, retention, and lifetime value.

One of the key insights from the results is that a small proportion of high-value customers contributed disproportionately to overall revenue. This finding aligns with the well-documented Pareto principle in marketing, where 20% of customers often account for 80% of sales. By identifying these customers through RFM clustering, the system offers businesses a strategy to focus promotional campaigns, personalized offers, and loyalty programs on the segments that matter most. Conversely, inactive and occasional buyersprovide an opportunity for re-engagement campaigns. Recommending trending or discounted products to such customers can potentially rekindle interest and increase transaction frequency.

The evaluation metrics further emphasized the suitability of K-Means clustering in segmenting retail customers. With a silhouette score of 0.67, K-Means outperformed Agglomerative Clustering and DBSCAN, confirming its ability to generate compact and well-separated clusters in this context. While DBSCAN's ability to detect noise was useful in flagging outliers, its relatively lower performance suggested that density-based clustering may not always capture the broader customer distribution effectively in transactional datasets. The comparison also illustrates that no single algorithm is universally superior; instead, the choice depends on dataset structure, noise levels, and the interpretability requirements of the application.

The integration of clustering with recommendation logic provided tangible value in practice. The mapping of product suggestions to cluster profiles demonstrated how machine learning can move beyond descriptive analytics toward actionable intelligence. For example, recommending premium décor products to high-value customers or offering promotional bundles to occasional buyers represents a pragmatic application of data-driven personalization. These strategies not only improve conversion rates but also enhance customer satisfaction by aligning suggestions with behavioral patterns.

Equally significant is the deployment of the recommendation system through a Streamlit-based interactive dashboard. The visual representation of customer clusters, sales trends, and recommendation outputs makes the system accessible to non-technical business stakeholders. Decision-makers can interact with the results, explore customer groups dynamically, and gain actionable insights without requiring advanced technical expertise. This usability aspect underscores the importance of designing machine learning solutions that are not only accurate but also interpretable and practical in real-world settings.

Despite the promising results, certain challenges and limitations were observed. The dataset, while rich, represents a single year of retail transactions, which may not fully capture evolving customer behaviors over longer periods. The reliance on RFM metrics, although powerful, does not consider additional features such as product categories, browsing history, or demographic information that could further refine segmentation. Moreover, the evaluation was based on offline metrics, and the system's effectiveness in live deployment scenarios—where user responses to recommendations may vary—remains an open question.

In conclusion, the discussion highlights that the proposed system successfully integrates clustering, recommendation, and visualization into a cohesive framework that addresses real-world challenges in e-commerce personalization. The findings validate the effectiveness of RFM-based clustering for customer segmentation while also emphasizing the importance of deployment through interactive dashboards for practical adoption. Future extensions could incorporate hybrid recommendation strategies, real-time transaction streams, and contextual data to further enhance system performance and adaptability.

V.CONCLUSION

The development and evaluation of the *Product Recommendation System* demonstrated the effectiveness of combining customer segmentation with recommendation logic to enhance e-commerce personalization. By leveraging RFM metrics and clustering algorithms, the system successfully identified distinct customer groups, each with unique purchasing behaviors. The results confirmed that a small segment of high-value customers contributes significantly to overall revenue, emphasizing the need for businesses to focus resources on retention and loyalty strategies for this group. At the same time, the identification of inactive and occasional customers created opportunities for targeted re-engagement through customized product recommendations and promotional offers.

The analysis of clustering algorithms revealed that K-Means provided the most effective segmentation, with the highest silhouette score and clear group distinctions. This confirmed its suitability for datasets with well-distributed transactional patterns, while also highlighting the limitations of other clustering methods such as DBSCAN in this context. The recommendation outcomes further illustrated the practical utility of aligning product suggestions with customer profiles, thereby enhancing the likelihood of conversions and improving overall customer experience.

A significant strength of the proposed system lies in its deployment through a Streamlit-based interactive dashboard. The dashboard not only enabled real-time visualization of clusters and recommendations but also made the results interpretable and actionable for business stakeholders without technical expertise. This user-friendly aspect underlines the importance of bridging the gap between machine learning outputs and decision-making processes in real-world applications.

However, the study also identified areas for improvement. The reliance on RFM features, while effective, excludes other potential behavioral and demographic data that could provide more comprehensive insights into customer preferences. Additionally, the offline evaluation of recommendations leaves room for future work to incorporate live A/B testing and real-time feedback mechanisms.

Expanding the dataset to cover longer timeframes and integrating hybrid recommendation models could further enhance accuracy and adaptability.

In conclusion, this research established a practical and scalable framework for customer segmentation and product recommendation in e-commerce. The integration of clustering techniques with recommendation logic and interactive visualization tools represents a significant step toward more personalized, data-driven customer engagement. With further enhancements and real-world deployment, such systems hold the potential to transform how businesses understand their customers and deliver tailored shopping experiences, ultimately driving sustained growth and customer satisfaction.

References

1. J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, Jul. 2013.
2. X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence*, vol. 2009, pp. 1–19, 2009.
3. A. Aggarwal, *Recommender Systems: The Textbook*. Springer, 2016.
4. P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, 2nd ed., Pearson Education, 2019.
5. D. M. Hawkins, "The problem of overfitting," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 1, pp. 1–12, Jan. 2004.
6. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., Springer, 2009.
7. M. Steinbach, G. Karypis, and V. Kumar, "A comparison of document clustering techniques," *Proc. KDD Workshop on Text Mining*, vol. 400, no. 1, pp. 525–526, 2000.
8. D. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
9. R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, no. 2–3, pp. 271–274, 1998.
10. J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Elsevier, 2011.
11. UCI Machine Learning Repository, "Online Retail Dataset," [Online]. Available: https://archive.ics.uci.edu/ml/datasets/online+retail [Accessed: Jun. 20, 2025].
12. Y. Zhao and G. Karypis, "Criterion functions for document clustering: Experiments and analysis," University of Minnesota, Department of Computer Science, Tech. Rep. 01-40, 2001.
13. S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proc. 25th Conf. on Uncertainty in Artificial Intelligence (UAI)*, 2009, pp. 452–461.
14. L. Rokach and B. Shapira, "Recommender systems: Introduction and challenges," in *Recommender Systems Handbook*, Springer, 2015, pp. 1–34.
15. F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds., *Recommender Systems Handbook*. Springer, 2011.