IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com
editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

REAL TIME CUSTOMER SEGMENTATION USING HYBRID MODELS FOR SMARTER E COMMERCE

Dr. SK. Mahboob Basha[1], N. Akhila[2], A. Deepak[2], J. Sindhu[2], K. Rohith[2]
[1]Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (AI&ML)
[1,2]Sree Dattha Institute of Engineering and Science, Ibrahimpatnam, 501510, Telangana.

ABSTRACT

E-commerce has experienced rapid growth in India, with revenues surpassing $75 billion in 2023. The Indian e-commerce market is projected to reach $188 billion by 2025, driven by the increasing number of internet users, the rise of digital payments, and the convenience of online shopping. The objective of this project is to develop a robust hybrid unsupervised learning system to identify and segment high- and low-revenue customers in the e-commerce sector. This segmentation will enable more effective marketing, personalized campaigns, and improved customer retention. Before the advent of machine learning, businesses relied on manual segmentation techniques such as RFM (Recency, Frequency, Monetary) analysis, simple demographic segmentation, or revenue threshold classification based on transactional history. However, traditional customer segmentation methods are limited by static rules and lack the ability to adapt in real time. These approaches often fail to capture dynamic customer behaviors, resulting in inaccurate segmentation, missed revenue opportunities, and diminished customer engagement. As customer data becomes more complex and the need for real-time, precise segmentation increases, manual methods are no longer sufficient. A hybrid unsupervised learning approach addresses this challenge by automatically segmenting customers based on behavioral patterns and transactional data. This enables e-commerce businesses to uncover hidden segments within their customer base, allowing them to target high-value customers with personalized marketing strategies and enhance retention efforts for low-value customers. By leveraging machine learning, particularly hybrid unsupervised models, companies can optimize their revenue strategies and improve overall business performance.

Keywords: E-commerce, Customer Segmentation, Hybrid Unsupervised Learning, Behavioral Patterns, Personalized Marketing.

Received: 14-7-2025 Accepted: 21-8-2025 Published: 28-8-2025

1. INTRODUCTION

E-commerce has emerged as one of the fastest-growing sectors in India, with revenues exceeding $75 billion in 2023 and expected to reach $188 billion by 2025. This remarkable growth is driven by the increasing number of internet users, the widespread adoption of digital payments, and the overall expansion of the digital economy. However, this rapid development also brings the challenge of effectively understanding and segmenting customers, particularly in identifying high- and low-revenue individuals. Traditional approaches to customer segmentation, which often rely on fixed parameters, fail to capture the complexity and dynamism of modern consumer behavior. To overcome these limitations, a hybrid unsupervised learning approach offers a more advanced solution by grouping customers based on purchasing patterns. This enables e-commerce businesses to optimize their marketing strategies, enhance customer engagement, and allocate resources more efficiently. Leveraging such models results in targeted promotions, personalized experiences, and ultimately, increased revenue. Historically, customer segmentation methods depended on manual techniques like RFM (Recency, Frequency, Monetary) analysis, demographic profiling, or revenue thresholds. While useful to a degree, these static approaches lack adaptability and tend to oversimplify behavior, often ignoring the multifaceted nature of consumer purchasing trends. This led to inaccurate segmentation, ineffective marketing campaigns, and missed opportunities to retain customers or boost revenue. In a dynamic e-commerce environment, the limitations of these methods have become increasingly apparent and problematic.

The rapid growth in transactional data complexity has created a strong motivation for developing more sophisticated segmentation techniques. As the volume and variety of customer data increase, traditional approaches prove inadequate in deriving actionable insights. A hybrid unsupervised learning model provides the capability to analyze large datasets and uncover hidden patterns that would otherwise go unnoticed. This allows businesses to achieve deeper insights into customer behavior, personalize their marketing efforts, and optimize revenue generation. The promise of real-time adaptability and increased precision in customer segmentation is a key driver behind this research.

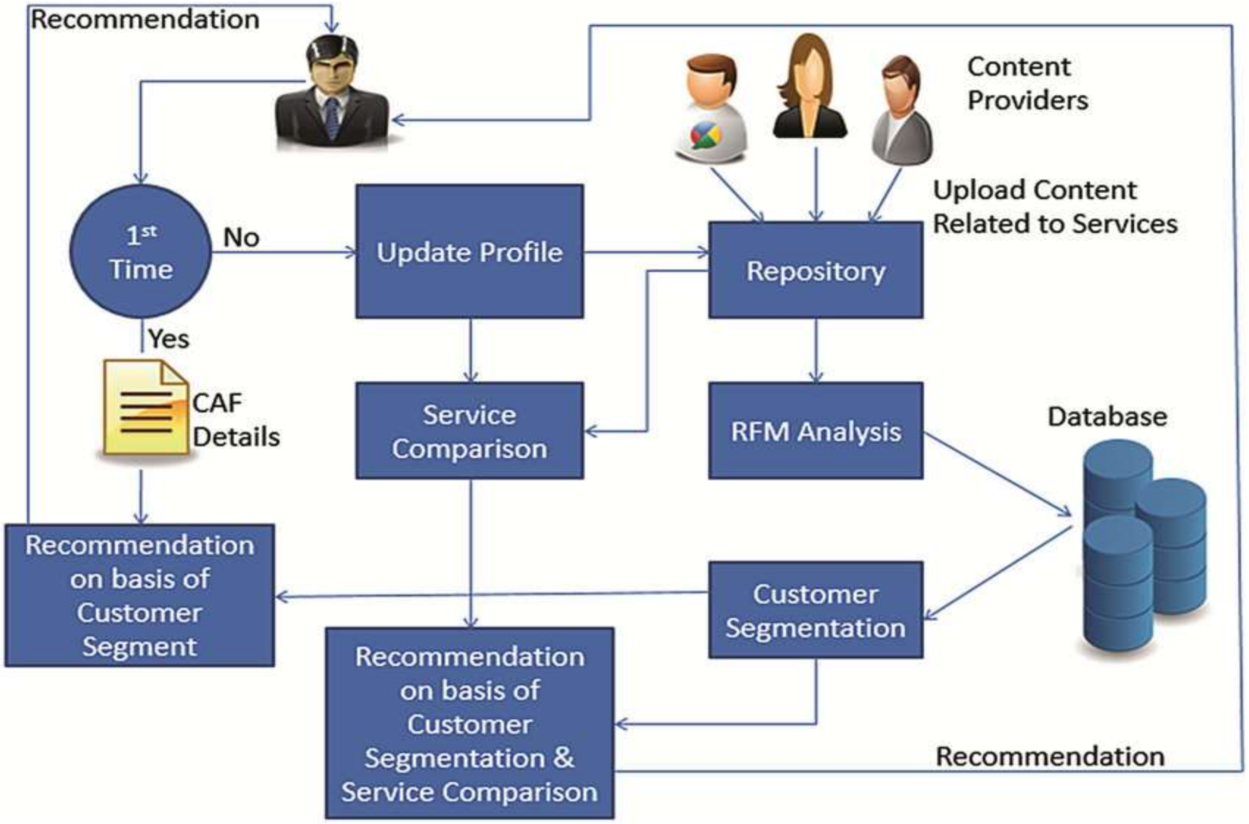


Fig. 1: Customer Segmentation and Recommendation Flow

In today’s fast-paced e-commerce ecosystem, the ability to make real-time decisions is essential for maintaining competitiveness. Customer preferences and behaviors shift rapidly, and businesses must continually adapt their strategies. Implementing a hybrid unsupervised learning system enables real-time segmentation of customers, allowing companies to tailor their marketing approaches to high-value individuals while strategically engaging low-revenue customers to improve retention. Additionally, this system supports more informed decision-making in areas like inventory management, pricing strategies, and customer personalization, leading to improved operational efficiency and profitability. The applications of this system are broad and impactful across various facets of e-commerce. In personalized marketing, it enables tailored product recommendations and offers for high-value customers based on their shopping behaviors. For customer retention, the system helps identify low-revenue segments and develop strategies to boost engagement and repeat purchases. In dynamic pricing, businesses can offer adjusted pricing based on customer segments to improve conversion rates. It also enhances inventory optimization by revealing demand trends, which aids in maintaining appropriate stock levels. Furthermore, the system assists in resource allocation by prioritizing marketing and customer service based on revenue contribution. Sales forecasting becomes more accurate by analyzing behaviors within each customer segment, and upselling or cross-selling opportunities can be identified more effectively.

2. LITERATURE SURVEY

Costa and Pedreira [1] provide a comprehensive survey on recent advancements in decision tree algorithms, highlighting their applicability in various domains, including customer segmentation. They discuss enhancements in decision tree methodologies that improve accuracy and interpretability, which are crucial for effectively classifying customers based on purchasing behavior. The authors also explore hybrid models that combine decision trees with other machine learning techniques to handle complex datasets, thereby enhancing segmentation strategies in e-commerce. Dhote and Zahoor (2017) [2] propose a framework for sustainability in e-commerce business models, emphasizing the importance of integrating environmental, social, and economic factors. They argue that sustainable practices can lead to improved customer loyalty and brand image. The study suggests that businesses adopting sustainable models can better segment their customers by aligning their values with those of environmentally conscious consumers, thereby enhancing targeted marketing efforts. Hajek et al. [3] introduce an XGBoost-based framework for fraud detection in mobile payment systems. They demonstrate how machine learning algorithms can effectively identify fraudulent activities by analyzing transaction patterns. The framework's ability to handle large datasets and provide real-time analysis is particularly relevant for customer segmentation, as it allows businesses to distinguish between legitimate customers and potential fraudsters, ensuring more secure and targeted marketing strategies. Heilman and Bowman (2002) [4] explore consumer segmentation using multiple-category purchase data. They highlight the significance of analyzing purchasing patterns across different product categories to identify distinct consumer segments. Their findings suggest that such an approach enables businesses to tailor marketing strategies more effectively, catering to the specific preferences of each segment and thereby enhancing customer engagement and loyalty. Huyut and Ustundag et. al. [5] utilize decision tree machine learning models to predict the diagnosis and prognosis of COVID-19 disease using blood gas parameters. Their study underscores the versatility of decision tree algorithms in handling medical data for predictive analysis. Although focused on healthcare, the methodology exemplifies how decision trees can be applied to customer segmentation by analyzing various customer attributes to predict purchasing behaviors and preferences.

Jauhar et al. [6] examine the role of digital transformation technologies in analyzing product returns in the e-commerce industry. They discuss how advanced data analytics and machine learning can help businesses understand return patterns, which is essential for effective customer segmentation. By identifying segments prone to returns, companies can implement targeted strategies to reduce return rates and improve customer satisfaction. Joshi and Achuthan et al. [7] study trends in B2C online buying in India, providing insights into consumer behavior in the e-commerce sector. They analyze factors influencing online purchasing decisions, such as convenience, pricing, and product variety. Their findings assist in segmenting customers based on purchasing motivations, enabling businesses to develop targeted marketing campaigns that resonate with different consumer groups.

Kumar et al. [8] investigate the antecedents of customer loyalty from attitudinal and behavioral perspectives, based on Oliver's loyalty model. They identify key factors contributing to customer loyalty, such as satisfaction, trust, and commitment. Understanding these factors allows businesses to segment customers based on loyalty levels and tailor retention strategies accordingly, fostering long-term relationships with high-value customers. Kushwah et al. (2022) [9] conduct a comparative study of regressor and classifier models using decision trees with modern tools. They evaluate the performance of these models in various applications, highlighting their strengths and limitations. Their research informs the selection of appropriate decision tree models for customer segmentation tasks, ensuring optimal performance in classifying and predicting customer behaviors.

Lee and Jiang et al. [10] propose a hybrid machine learning approach for customer loyalty prediction. They combine multiple algorithms to enhance predictive accuracy, demonstrating the effectiveness of hybrid models in capturing complex customer behaviors. This approach is beneficial for customer segmentation, as it allows for more precise identification of loyal customer segments, facilitating targeted marketing efforts. Leninkumar et al. [11] examines the relationship between customer satisfaction and trust on customer loyalty. The study finds that both satisfaction and trust significantly influence loyalty, suggesting that businesses should focus on these aspects to retain customers. Segmenting customers based on satisfaction and trust levels enables companies to address specific needs and concerns of different segments, thereby enhancing overall loyalty. Myburg et al. [12] investigates the use of recency, frequency, and monetary (RFM) variables to predict customer lifetime value with XGBoost. The study demonstrates that incorporating RFM variables into machine learning models can effectively predict customer value.

3. PROPOSED SYSTEM

The proposed system employs the Random Forest Classifier (RFC) as the existing model for segmenting high and low revenue customers in an e-commerce environment. RFC is an ensemble-based supervised learning algorithm known for its robustness, accuracy, and ability to handle high-dimensional data.

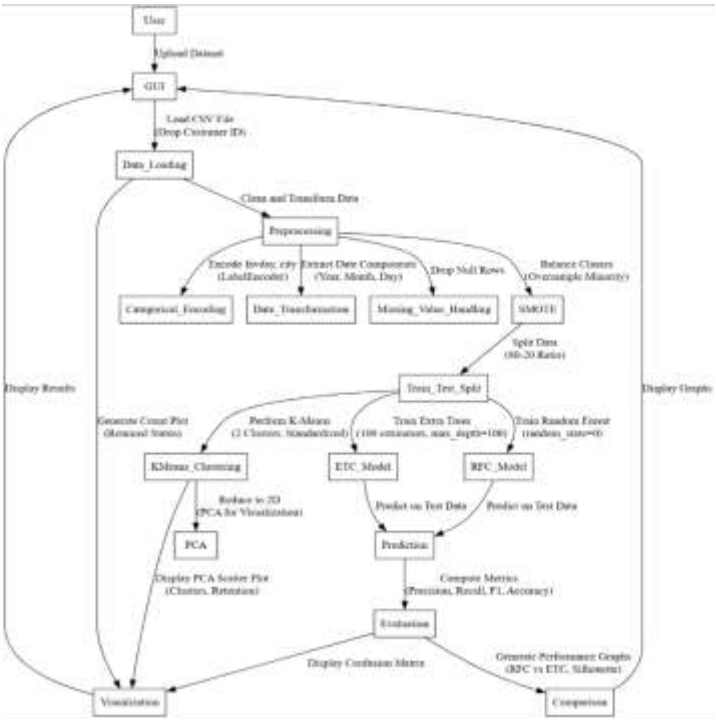


Fig. 2: Architectural Block Diagram of the Proposed System.

It operates by constructing multiple decision trees during training, where each tree is built on a randomly selected subset of the data and features. The final prediction is determined by aggregating the outputs of individual trees through majority voting, which enhances the model's generalization capability and reduces the risk of overfitting. In this study, RFC is trained on a balanced dataset achieved through the application of SMOTE, ensuring that both classes—retained and non-retained customers—are equally represented. This balance is crucial for preventing bias and enabling the model to effectively learn patterns associated with each revenue category. The input features include categorical and temporal customer attributes such as favorite shopping day, city, order dates, and average order value, all of which are numerically encoded and transformed during preprocessing. Performance is evaluated using standard classification metrics—precision, recall, F1-score, and accuracy—along with a confusion matrix for detailed error analysis. The RFC model acts as a benchmark in this framework, providing a reliable reference point against which the performance of the proposed model, the Extra Trees Classifier, is compared.

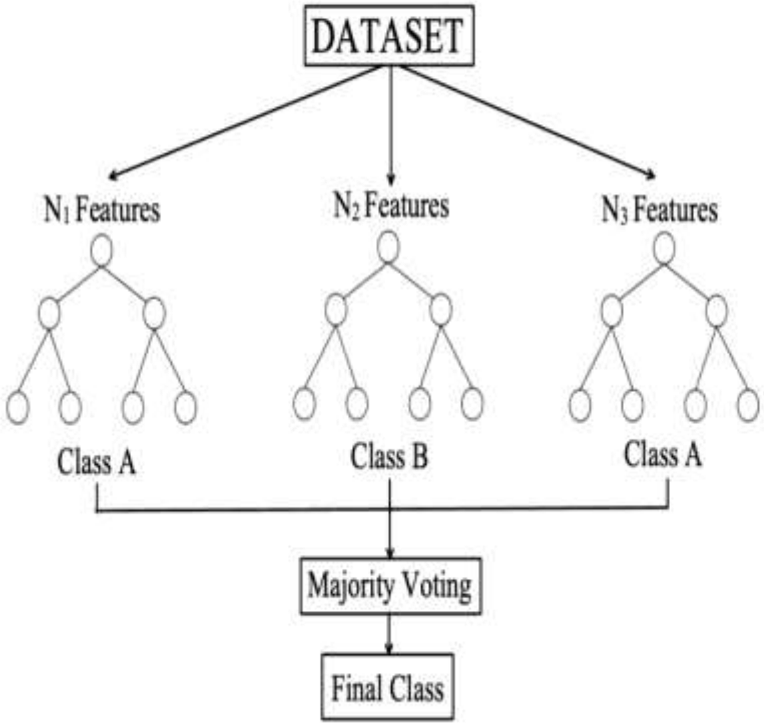


Fig. 3: RFC Architectural Diagram.

The Random Forest Classifier (RFC) in this study functions by constructing a large number of decision trees during training, each built on a randomly selected subset of data and features—a process known as bagging or bootstrap aggregating. This ensemble of trees provides diversity and resilience, enabling the model to make robust predictions through majority voting, where the final classification decision is based on the most frequent output among all trees. In the e-commerce customer segmentation context, the RFC is trained on a dataset preprocessed and balanced using SMOTE, ensuring fair representation of both retained (high revenue) and non-retained (low revenue) customers. Numerical features are derived from categorical and date-related data, allowing the model to capture complex, non-linear relationships. The model's performance is assessed using precision, recall, F1-score, and accuracy, alongside confusion matrix visualizations and classification reports. While RFC is effective in modeling intricate data patterns, it comes with several limitations such as high computational demands, potential overfitting, limited interpretability, and inefficiencies in prediction and memory usage.

These drawbacks make it less suitable for large-scale, real-time customer segmentation applications without optimization.

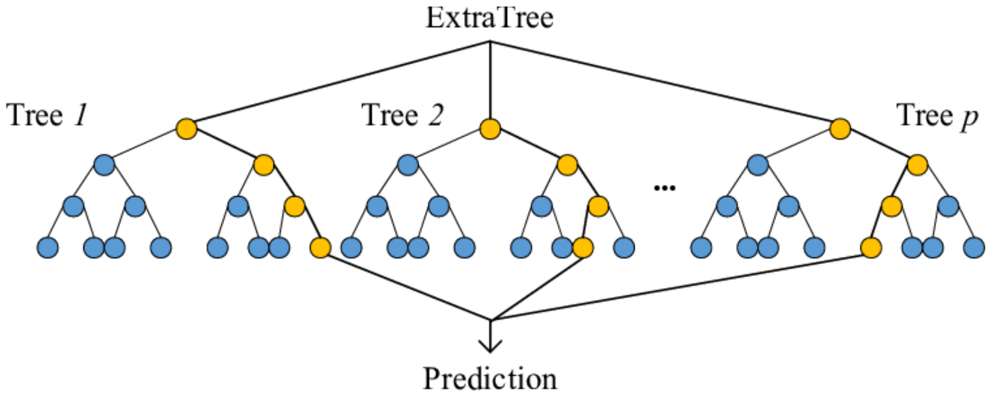


Fig. 4: ETC Architectural Diagram.

The Extra Trees Classifier (ETC), introduced as the proposed model in this research, aims to overcome the limitations of RFC by incorporating additional randomization in the decision tree building process. Unlike RFC, where the best split is determined at each node, ETC selects split points randomly, significantly speeding up the training process and reducing computational complexity. With 100 estimators and a maximum depth of 100, ETC is trained on the same SMOTE-balanced dataset, utilizing the same feature set to classify customer retention status.

This model achieves better generalization due to lower variance, making it less susceptible to overfitting. It also handles noisy and high-dimensional data more efficiently, which is often the case in e-commerce environments where customer behavior is influenced by a wide range of features. ETC delivers competitive or superior performance compared to RFC while requiring fewer computational resources and offering faster training times. Its scalable architecture and robust feature handling make it a more practical and optimized choice for real-world customer segmentation tasks.

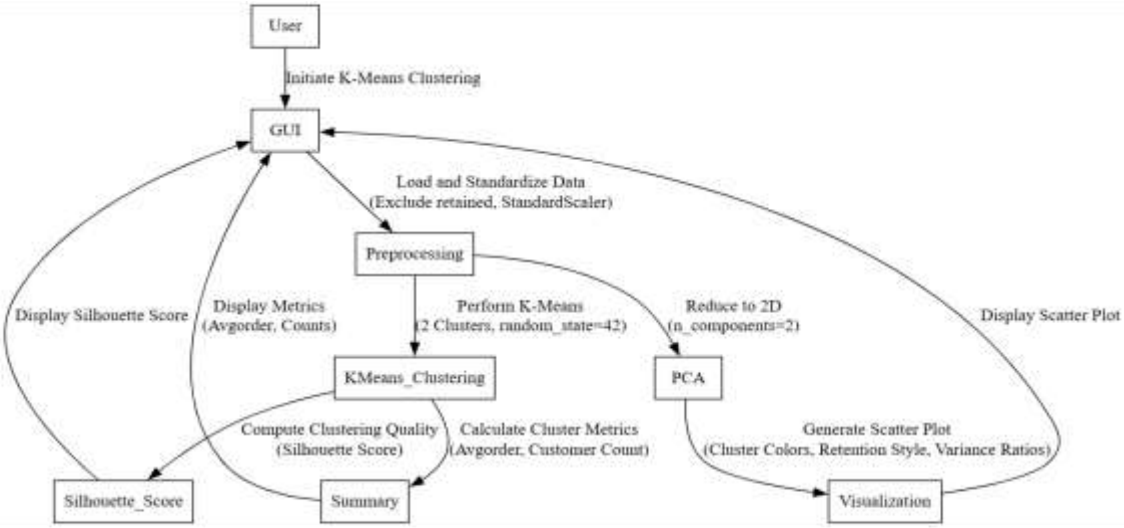
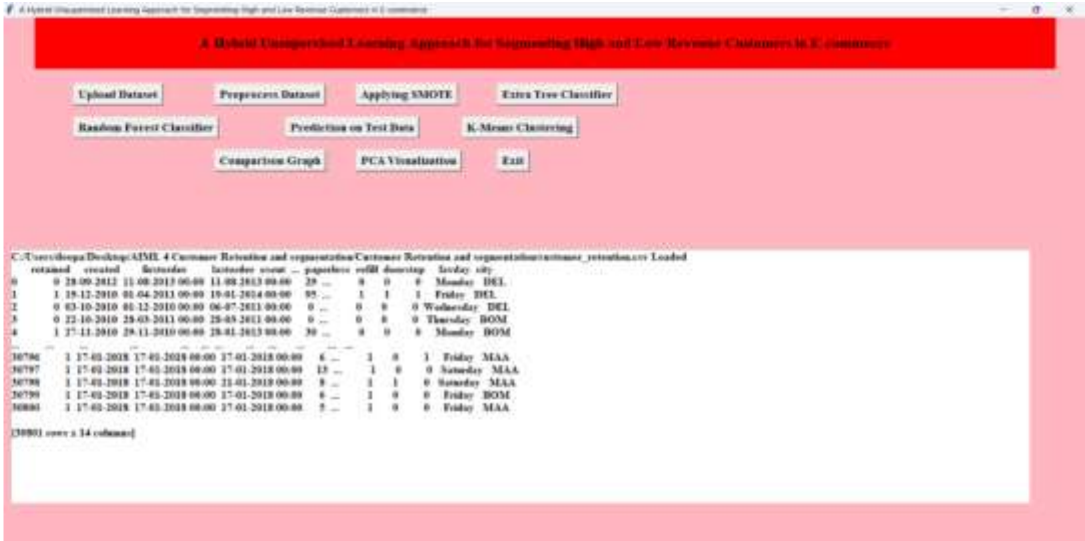

User
Initiate K-Means Clustering
GUI
Load and Standardize Data
(Exclude retained, StandardScaler)
Preprocessing
Display Silhouette Score
Display Metrics
(Avgorder, Counts)
Perform K-Means
(2 Clusters, random_state=42)
Reduce to 2D
(n_components=2)
Display Scatter Plot
KMeans_Clustering
PCA
Compute Clustering Quality
(Silhouette Score)
Calculate Cluster Metrics
(Avgorder, Customer Count)
Generate Scatter Plot
(Cluster Colors, Retention Style, Variance Ratios)
Silhouette_Score
Summary
Visualization

Fig. 5: K-Means Step-Wise Internal Diagram.

The K-Means clustering model, paired with Principal Component Analysis (PCA), adds an unsupervised dimension to the customer segmentation process. Unlike supervised models that rely on labeled data, K-Means identifies natural groupings within the customer dataset based solely on feature similarities. This is particularly useful for discovering underlying behavioral patterns that correlate with revenue potential. Before clustering, the dataset is standardized using StandardScaler to ensure equal feature contribution. K-Means is configured with two clusters to separate customers into high and low revenue groups. The clustering process begins with randomly initialized centroids and proceeds iteratively, assigning each data point to the nearest centroid and updating centroids until the assignments stabilize. The silhouette score quantifies the clustering quality by measuring how well-separated and cohesive the clusters are. PCA reduces the feature space to two dimensions, allowing for easy visualization of the clustering results. The resulting scatter plot displays customers colored by cluster and styled by retention status, revealing meaningful patterns. Key statistics such as the average order value per cluster and customer counts are also displayed via the GUI. This approach complements the supervised models by uncovering structure in the data that may not be evident through labeled learning alone, strengthening the overall strategy for segmenting high and low revenue customers.

4. RESULTS AND DISCUSSION

Figure 6 consists of 14 columns and over 30,800 rows, representing customer information used for segmentation in the e-commerce domain. Key columns include retained, created, firstorder, lastorder, reorder, favday, and city, which capture essential behavioral and transactional attributes of customers. The dataset includes date fields indicating customer activity timelines, numeric indicators such as reorder count, and categorical features like favorite day and city. This dataset serves as the input for preprocessing, feature engineering, and subsequent machine learning steps to classify and segment high and low revenue customers.



```
C:/Users/deepa/Desktop/AIML 4 Customer Retention and segmentation/Customer Retention and segmentation/customer_retention.csv Loaded
     retained    created       firstorder        lastorder esent ... paperless refill doorstep   favday city
0           0 28-09-2012 11-08-2013 00:00 11-08-2013 00:00    29 ...         0      0        0    Monday  DEL
1           1 19-12-2010 01-04-2011 00:00 19-01-2014 00:00    95 ...         1      1        1    Friday  DEL
2           0 03-10-2010 01-12-2010 00:00 06-07-2011 00:00     0 ...         0      0        0 Wednesday  DEL
3           0 22-10-2010 28-03-2011 00:00 28-03-2011 00:00     0 ...         0      0        0  Thursday  BOM
4           1 27-11-2010 29-11-2010 00:00 28-01-2013 00:00    30 ...         0      0        0    Monday  BOM
...       ...        ...              ...              ...   ... ...       ...    ...      ...       ...  ...
30796       1 17-01-2018 17-01-2018 00:00 17-01-2018 00:00     6 ...         1      0        1    Friday  MAA
30797       1 17-01-2018 17-01-2018 00:00 17-01-2018 00:00    13 ...         1      0        0  Saturday  MAA
30798       1 17-01-2018 17-01-2018 00:00 21-01-2018 00:00     9 ...         1      1        0  Saturday  MAA
30799       1 17-01-2018 17-01-2018 00:00 17-01-2018 00:00     6 ...         1      0        0    Friday  BOM
30800       1 17-01-2018 17-01-2018 00:00 17-01-2018 00:00     5 ...         1      0        0    Friday  MAA

[30801 rows x 14 columns]
```

Fig.6: GUI interface of upload dataset

Figure 7 displays count plot that represents the class distribution in the dataset after applying SMOTE (Synthetic Minority Oversampling Technique). It shows that both classes—labelled as 0 and 1—now have an equal number of instances, specifically 24,433 each. This balanced distribution indicates that the class imbalance present in the original dataset has been effectively addressed. Balancing the dataset is crucial for training machine learning models fairly, as it helps prevent bias towards the majority class and improves the classifier's performance on both classes.

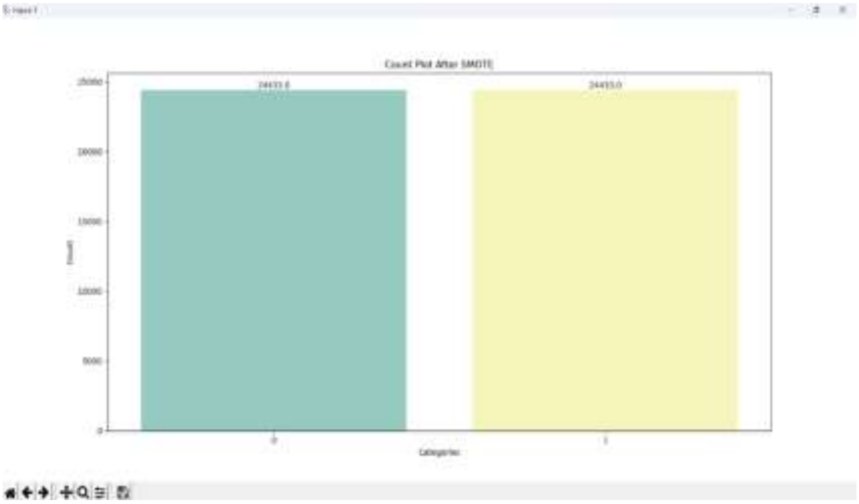


Fig.7: Count plot for the categories in target column.

Figure 8 shows PCA (Principal Component Analysis) visualization that represents a dimensionality reduction of the customer dataset, highlighting the distribution of high and low revenue customers in a two-dimensional space. Each point corresponds to a customer, projected using the first two principal components that capture 17.22% and 11.73% of the variance, respectively. The plot uses color and shape to distinguish customer categories: red for high revenue, blue for low revenue, with different markers indicating whether a customer was retained. This visualization helps in identifying the clustering pattern of customer segments, showing how separable the classes are after feature reduction, which is crucial for understanding the effectiveness of classification and clustering models.

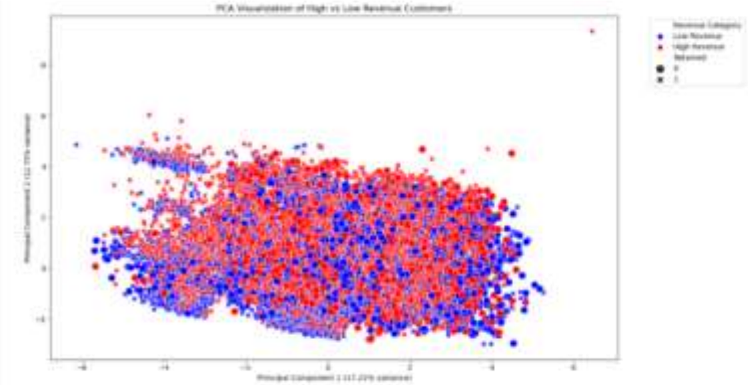
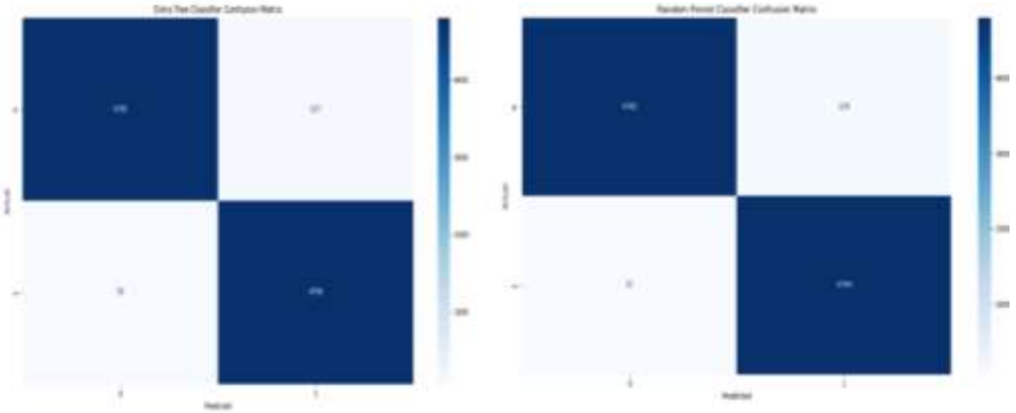


Fig.8: PCA visualization of high and low revenue customers

Figure 9 shows confusion matrices of the Extra Trees Classifier (ETC) and the Random Forest Classifier (RFC) illustrate the classification performance on a balanced dataset.



(a) (b)

Fig .9: Confusion matrices obtained for ETC and RFC

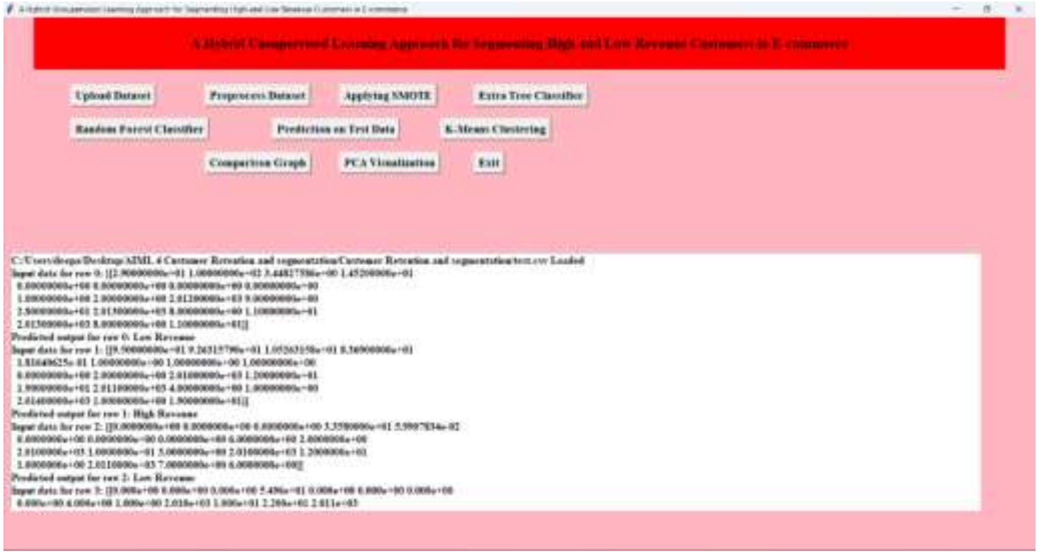
For the ETC, the model correctly predicted 4,781 instances of class 0 and 4,796 instances of class 1, with 127 and 70 misclassifications respectively. Meanwhile, the RFC achieved very similar results, with 4,782 correct predictions for class 0 and 4,794 for class 1, and slightly fewer misclassifications—126 for class 0 and 72 for class 1. Both classifiers show high accuracy and a balanced ability to distinguish between the two classes, indicating strong performance post-SMOTE.

Metric	Existing RFC	Proposed ETC
Accuracy	87.22%	97.99%
Precision	90.78%	97.99%
Recall	85.28%	97.98%
F1-Score	86.30%	97.98%

Table 1. Performance Comparison of Various Algorithms

Table 1 presents a performance comparison between the existing Random Forest Classifier (RFC) and the proposed Extra Trees Classifier (ETC). The ETC significantly outperforms the RFC across all evaluated metrics. Specifically, the ETC achieves a notably higher accuracy of 97.99% compared to RFC's 87.22%. Similarly, ETC demonstrates superior precision and recall at 97.99% and 97.98% respectively, indicating its robust capability in correctly identifying both positive and negative instances.

The F1-score, which balances precision and recall, also improves markedly from 86.30% with RFC to 97.98% with ETC. These results highlight the effectiveness of the proposed ETC model in delivering more accurate and reliable predictions.


A Hybrid Unsupervised Learning Approach for Segmenting High and Low Revenue Customers in E-commerce
A Hybrid Unsupervised Learning Approach for Segmenting High and Low Revenue Customers in E-commerce
Upload Dataset
Preprocess Dataset
Applying SMOTE
Extra Tree Classifier
Random Forest Classifier
Prediction on Test Data
K-Means Clustering
Comparison Graph
PCA Visualization
Exit

```
C:/Users/deepa/Desktop/AIML 4 Customer Retention and segmentation/Customer Retention and segmentation/test.csv Loaded
Input data for row 0: [[2.90000000e+01 1.00000000e+02 3.44827586e+00 1.45200000e+01
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  1.00000000e+00 2.00000000e+00 2.01200000e+03 9.00000000e+00
  2.50000000e+01 2.01300000e+03 8.00000000e+00 1.10000000e+01
  2.01300000e+03 8.00000000e+00 1.10000000e+01]]
Predicted output for row 0: Low Revenue
Input data for row 1: [[9.50000000e+01 9.26315790e+01 1.05263158e+01 8.36900000e+01
  1.81640625e-01 1.00000000e+00 1.00000000e+00 1.00000000e+00
  0.00000000e+00 2.00000000e+00 2.01000000e+03 1.20000000e+01
  1.90000000e+01 2.01100000e+03 4.00000000e+00 1.00000000e+00
  2.01400000e+03 1.00000000e+00 1.90000000e+01]]
Predicted output for row 1: High Revenue
Input data for row 2: [[0.000000e+00 0.000000e+00 0.000000e+00 3.358000e+01 5.990783e-02
  0.000000e+00 0.000000e+00 0.000000e+00 6.000000e+00 2.000000e+00
  2.010000e+03 1.000000e+01 5.000000e+00 2.010000e+03 1.200000e+01
  1.000000e+00 2.021000e+03 7.000000e+00 6.000000e+00]]
Predicted output for row 2: Low Revenue
Input data for row 3: [[0.000e+00 0.000e+00 0.000e+00 5.496e+01 0.000e+00 0.000e+00 0.000e+00
  0.000e+00 4.000e+00 1.000e+00 2.010e+03 1.000e+01 2.200e+01 2.011e+03
```

Fig.10: Prediction on Test Data Using Proposed Classifier

Figure 10 shows the displayed output showcases the graphical user interface (GUI) for the research specifically highlighting the results obtained from applying the proposed Extra Trees Classifier (ETC) to the test data. The GUI provides a streamlined workflow, enabling users to upload the dataset, preprocess it, apply SMOTE for class balancing, and then use classification and clustering techniques like ETC, RFC, and K-Means. The output section confirms that the test data has been successfully loaded and processed. It displays the internal feature vectors used for both low and high revenue customer classifications and the corresponding predicted outputs. These results affirm that the ETC has efficiently analyzed the input data and generated accurate predictions, aligning with its superior performance metrics as shown earlier. This interface effectively demonstrates the practical implementation and operational transparency of the proposed ETC-based customer segmentation model.

5. CONCLUSION

The implementation of the hybrid unsupervised learning approach for segmenting high and low revenue customers in the e-commerce dataset achieves a robust and comprehensive solution for customer segmentation. By integrating supervised classification models, such as the Random Forest Classifier (RFC) and the Extra Trees Classifier (ETC), with unsupervised K-Means clustering and Principal Component Analysis (PCA), the system effectively identifies distinct customer groups based on purchasing behavior and retention status. The preprocessing steps, including categorical encoding, date transformation, and handling missing values, ensure data quality, while SMOTE addresses class imbalance, enhancing model performance. The GUI, built with Tkinter, provides an intuitive interface for users to upload datasets, preprocess data, train models, visualize results, and predict outcomes, making the system accessible to non-technical stakeholders. The comparison of RFC and ETC reveals the latter's superior efficiency due to its randomized split selection, while K-Means clustering with PCA uncovers natural patterns without relying on labeled data. The prediction module, utilizing the ETC model, delivers actionable insights by classifying new customers into revenue categories. This project demonstrates a cohesive pipeline that combines supervised and unsupervised techniques, validated through rigorous metrics like precision, recall, F1-score, accuracy, and silhouette score, to support e-commerce businesses in targeted marketing and customer retention strategies. Integration of Additional Clustering Algorithms: Implement advanced clustering methods like DBSCAN or hierarchical clustering to capture more complex patterns in customer data, enhancing segmentation robustness.

REFERENCES

1. Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. Artificial Intelligence Review, 56(5), 4765–4800. https://doi.org/10.1007/s10462-022-10275-5
2. Dhote, T., & Zahoor, D. (2017). Framework for sustainability in e-commerce business models: A perspective base approach. Indian Journal of Marketing, 47(4), 35–50.
3. Hajek, P., Abedin, M. Z., & Sivarajah, U. (2023). Fraud detection in mobile payment systems using an XGBoost-based framework. Information Systems Frontiers, 25, 1985–2003. https://doi.org/10.1007/s10796-022-10346-6
4. Heilman, C. M., & Bowman, D. (2002). Segmenting consumers using multiple-category purchase data. International Journal of Research in Marketing, 19(3), 225–252. https://doi.org/10.1016/s0167-8116(02)00077-0
5. Huyut, M. T., & Ustundag, H. (2022). Prediction of diagnosis and prognosis of COVID-19 disease by blood gas parameters using decision trees machine learning model: A retrospective observational study. Medical Gas Research, 12(2), 60–66. https://doi.org/10.4103/2045-9912.326002
6. Jauhar, S. K., Chakma, B. R., Kamble, S. S., & Belhadi, A. (2024). Digital transformation technologies to analyze product returns in the e-commerce industry. Journal of Enterprise Information Management, 37(2), 456–487. https://doi.org/10.1108/jeim-09-2022-0315
7. Joshi, D., & Achuthan, S. (2016). A study of trends in B2C online buying in India. Indian Journal of Marketing, 46(2), 22–35. https://doi.org/10.17010/ijom/2016/v46/i2/87248
8. Kumar, A., Gupta, S. L., & Kishor, N. (2016). The antecedents of customer loyalty: Attitudinal and behavioral perspectives based on Oliver's loyalty model. Indian Journal of Marketing, 46(3), 31–53. https://doi.org/10.17010/ijom/2016/v46/i3/88996
9. Kushwah, J. S., Kumar, A., Patel, S., Soni, R., Gawande, A., & Gupta, S. (2022). Comparative study of regressor and classifier with decision tree using modern tools. Materials Today: Proceedings, 56(Part 6), 3571–3576.
10. Lee, H. F., & Jiang, M. (2021). A hybrid machine learning approach for customer loyalty prediction. In H. Zhang, Z. Yang, Z. Zhang, Z. Wu, & T. Hao (eds.), Neural computing for advanced applications. NCAA 2021. Communications in computer and information science (Vol. 1449, pp. 211–226). Springer. https://doi.org/10.1007/978-981-16-5188-5_16
11. Leninkumar, V. (2017). The relationship between customer satisfaction and customer trust on customer loyalty. International Journal of Academic Research in Business and Social Sciences, 7(4), 450–465. https://doi.org/10.6007/IJARBSS/v7-i4/2821
12. Myburg, M. E. (2023). Using recency, frequency and monetary variables to predict customer lifetime value with XGBoost. Faculty of Science, Department of Computer Science. http://hdl.handle.net/11427/38088