Study and Comparative Analysis of Machine Learning Algorithms on Consumer Behaviour in Digital Marketplace

Shubham Agrawal
Information Technology
Thakur College of Engineering and Technology
Mumbai, India
shubhamagrawal5801@gmail.com

Dr. Neeta Patil
Information Technology
Thakur College of Engineering and Technology
Mumbai, India
neeta.patil@tcetmumbai.in

Dr. Rajesh Bansode
Information Technology
Thakur College of Engineering and Technology
Mumbai, India
rajesh.bansode@thakureducation.org

Abstract

The rapid expansion of e-commerce platforms has led to the generation of vast amounts of consumer transaction data, creating a need for intelligent recommendation systems to enhance user experience and drive sales. This study presents a comparative analysis of five widely used recommendation algorithms—K-Nearest Neighbors (KNN) based User-Based Collaborative Filtering, Item-Based Collaborative Filtering, Apriori, FP-Growth, and ECLAT—applied to a grocery purchase dataset.
The dataset consists of customer transaction records, which are preprocessed through data cleaning, transformation, and the construction of a user-item interaction matrix. Collaborative filtering techniques (KNN and Item-Based CF) are employed to generate personalized recommendations based on user similarity and item similarity, respectively. In contrast, association rule mining techniques (Apriori, FP-Growth, and ECLAT) are utilized to discover frequent itemsets and identify meaningful relationships between products.

The performance of these algorithms is evaluated using key metrics such as accuracy, precision, recall, and processing time to assess both recommendation quality and computational efficiency. Experimental results indicate that FP-Growth and ECLAT outperform Apriori in terms of processing speed while maintaining strong association discovery capabilities. Among collaborative filtering approaches, Item-Based Collaborative Filtering demonstrates improved scalability and consistency compared to User-Based KNN.

The findings highlight the strengths and limitations of each approach and emphasize the importance of selecting appropriate algorithms based on dataset characteristics and application requirements. This study contributes to the development of efficient and scalable recommendation systems, enabling e-commerce platforms to deliver personalized shopping experiences and optimize business outcomes.

Keywords — E-commerce, machine learning, consumer behaviour, Decision Tree, Random Forest, K-Means Clustering, predictive analytics, customer segmentation, marketing optimization.

I. INTRODUCTION

The rapid advancement of digital technologies has significantly transformed the e-commerce landscape, enabling businesses to leverage large volumes of data for strategic decision-making. With the increasing adoption of Artificial Intelligence (AI) and data-driven methodologies, organizations are now able to analyze consumer behaviour more effectively and deliver personalized experiences at scale (1), (2), (3). AI-powered systems have become integral to modern e-commerce platforms, enhancing marketing strategies, optimizing operations, and improving customer engagement (4), (5).

Understanding consumer behaviour is a critical component in the success of e-commerce systems. Customer purchasing decisions are influenced by multiple factors, including browsing history, product preferences, pricing strategies, and personalized recommendations (6), (7), (8). The ability to analyze these behavioural patterns allows businesses to predict future purchases, improve customer retention, and maximize revenue. However, the increasing volume, velocity, and variety of e-commerce data pose significant challenges in extracting meaningful and actionable insights efficiently (9), (10).

To address these challenges, recommender systems have emerged as a key application of machine learning in e-commerce. These systems aim to provide personalized product suggestions by analyzing historical user interactions and transactional data. Among the widely adopted approaches, collaborative filtering and association rule mining have shown significant effectiveness in generating recommendations (11), (12). Collaborative filtering techniques utilize similarities between users or items to predict user preferences, while association rule mining focuses on identifying frequent patterns and relationships among items in transactional datasets (13), (14).

Recent research highlights the growing importance of AI-driven recommendation systems in enhancing customer experience and influencing purchasing behaviour. Studies have demonstrated that personalized recommendations not only improve user satisfaction but also significantly impact conversion rates and business performance (15), (16), (17). Furthermore, advancements in AI and big data analytics have enabled the development of scalable and efficient recommendation models capable of handling large and dynamic datasets (18), (19).

Despite these advancements, different recommendation algorithms exhibit varying performance in terms of accuracy, scalability, and computational efficiency. Collaborative filtering methods are effective for personalization but may suffer from sparsity and scalability issues, whereas association rule mining techniques are efficient in discovering item relationships but can be computationally intensive depending on the algorithm used (20), (21), (22). Therefore, a comparative analysis of these techniques is essential to determine their suitability for real-world e-commerce applications.

In this context, the present study conducts a comparative analysis of five prominent recommendation algorithms—K-Nearest Neighbors (KNN) based User-Based Collaborative Filtering, Item-Based Collaborative Filtering, Apriori, FP-Growth, and ECLAT—using a grocery transaction dataset. The algorithms are evaluated based on key performance metrics such as accuracy, precision, recall, and processing time to assess both recommendation effectiveness and computational efficiency.

The objectives of this research are as follows:

1. To analyze consumer purchase behaviour using collaborative filtering and association rule mining techniques.
2. To compare the performance of different recommendation algorithms in terms of accuracy and efficiency.
3. To identify the most suitable algorithm for developing scalable and effective recommendation systems in e-commerce environments.

This study aims to contribute to the existing body of knowledge by providing a comprehensive evaluation of recommendation techniques and offering practical insights for businesses to enhance customer experience, optimize marketing strategies, and improve overall operational performance in digital marketplaces.

II. Literature survey

Recommender Systems in Retail

The rapid growth of e-commerce has led to increased adoption of Artificial Intelligence (AI) techniques for analyzing consumer behaviour and improving business strategies. Recent studies emphasize that AI-driven systems enhance customer engagement and enable personalized marketing by utilizing large-scale transactional data (1), (3). These advancements allow businesses to better understand user preferences and optimize decision-making processes in digital marketplaces (2).

Consumer behaviour in e-commerce is influenced by multiple factors such as personalization, user experience, and trust. Research indicates that AI-based recommendation systems significantly impact purchase decisions and customer satisfaction by delivering relevant product suggestions (6), (7). The ability to analyze historical data and predict future preferences has become a key factor in improving customer retention and sales performance (8).

Recommender systems have become an essential component of modern e-commerce platforms, enabling personalized product suggestions through various techniques. Among these, collaborative filtering methods are widely used to generate recommendations based on user-item interactions, with both user-based and item-based approaches improving accuracy and scalability in real-world applications (12), (14). Additionally, association rule mining techniques are employed to identify relationships between items in transactional datasets, supporting market basket analysis and enhancing cross-selling strategies (17), (18). Recent research focuses on improving the efficiency of these methods to handle large datasets and real-time recommendation scenarios.

Despite these advancements, selecting the most suitable recommendation technique remains challenging due to variations in accuracy, scalability, and computational efficiency. Hence, a comparative analysis of collaborative filtering and association rule mining approaches is essential to evaluate their effectiveness in practical e-commerce environments (21). This study addresses this need by comparing five algorithms—KNN, Item-Based Collaborative Filtering, Apriori, FP-Growth, and ECLAT—on a grocery dataset to identify the most efficient approach for personalized recommendations.

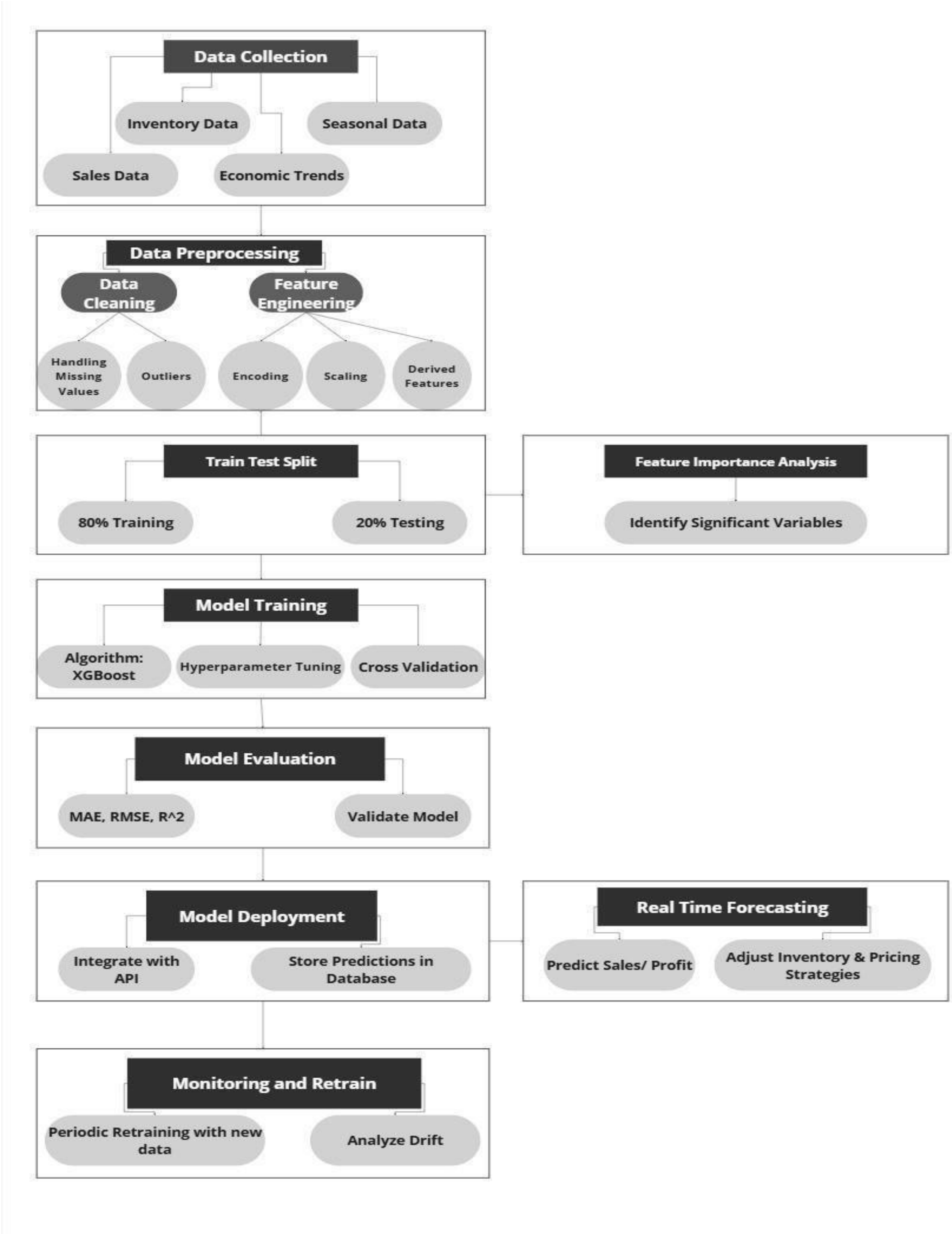


Fig. 1: Architecture for sales prediction

IV. METHODOLOGY

This research adopts a systematic approach to analyze grocery purchase data and compare the performance of different recommendation algorithms. The methodology consists of the following key steps, elaborated with greater detail:

1. **Data Acquisition & Preprocessing :**

The "Groceries_dataset.csv" is used, containing customer purchase data with 'Member_number', 'itemDescription', and 'Date'.
The 'Date' column is converted to datetime format for consistency. 'itemDescription' is label encoded for ML compatibility.
Items purchased 100+ times are retained to focus on frequently bought items. A sample of the dataset is used to speed up analysis while keeping it representative.

2. **Algorithms / Model Implementation**

KNN
Steps Involved in K-Nearest Neighbours (KNN) Algorithm :

Load the Data:
Load the dataset and perform necessary preprocessing (e.g., encoding categorical variables).
Create a user-item matrix: user_item_matrix = df_filtered.pivot_table(...)

Choose the Value of K:
This value (number of neighbors) is implicitly set when calling kneighbors():

distances, indices = knn.kneighbours(sample_user, n_neighbours=5)

Here, `n_neighbors=5` means K = 5.

Calculate Distance:
This is done automatically within the kneighbors() function.
The distance metric is specified when initializing the KNN model:

knn = NearestNeighbors(metric='cosine', algorithm='brute')

metric='cosine' uses cosine similarity to calculate distances between users in the user-item matrix.

Find Nearest Neighbours:
kneighbors() identifies the K nearest neighbors to the target user and stores their information in distances and indices.

Return the Result:
The code prints the indices of the neighbors, but in a real recommendation system, it would return a list of recommended items.

In addition to KNN, several other recommendation algorithms are employed in this study. Item-Based Collaborative Filtering identifies similarities between items using user-item interactions and recommends products similar to those previously purchased by the user. Association rule mining techniques such as Apriori, FP-Growth, and ECLAT are used to discover relationships between items in transactional data. Apriori generates frequent itemsets using support and confidence but is computationally intensive, while FP-Growth improves efficiency using a tree-based structure. ECLAT further enhances performance by utilizing a vertical data format and set intersection methods for faster frequent itemset generation.

3. Evaluation:

Accuracy: The accuracy of each algorithm is evaluated based on its ability to generate relevant recommendations and identify meaningful patterns in the dataset. For collaborative filtering methods such as KNN and Item-Based Collaborative Filtering, accuracy is assessed by analyzing how well the recommended items match user purchase behaviour. For association rule mining techniques including Apriori, FP-Growth, and ECLAT, accuracy is evaluated using measures such as support, confidence, and lift, which indicate the strength and reliability of item associations.

Processing Time: Processing time is measured to evaluate the computational efficiency of each algorithm. This is calculated by recording the start and end time of execution and determining the total time taken. Computational efficiency is an important factor in real-time recommendation systems, where faster algorithms provide quicker responses and improved user experience.

Visualization: Visualization techniques are used to better understand the dataset and interpret the results of the algorithms. Bar charts are used to represent the frequency of purchased items, histograms show the distribution of items per user, and heatmaps illustrate user-item interactions. These visualizations help in identifying patterns, trends, and overall performance of the models.

4. Comparison and Analysis:

The performance of the implemented algorithms is compared based on accuracy, processing time, and overall effectiveness in generating recommendations. Among the collaborative filtering techniques, KNN provides personalized recommendations by identifying similar users, while Item-Based Collaborative Filtering offers improved scalability by focusing on item similarity. Item-Based CF generally performs more consistently in larger datasets due to reduced computational complexity.

In the case of association rule mining techniques, Apriori, FP-Growth, and ECLAT demonstrate varying levels of efficiency. Apriori effectively identifies item associations but requires higher computational time due to repeated dataset scans. FP-Growth improves performance by utilizing a tree-based structure, resulting in faster execution and better scalability. ECLAT further enhances efficiency by using a vertical data format and set intersection methods, making it suitable for dense datasets.

Overall, FP-Growth and ECLAT provide better computational efficiency compared to Apriori, while maintaining strong association discovery capabilities. Collaborative filtering methods, particularly Item-Based CF, perform well in delivering personalized recommendations. The comparison highlights that the choice of algorithm depends on the specific requirements of the system, such as accuracy, scalability, and response time.

TABLE I: DATASET DESCRIPTION OF AI-DRIVEN MARKETING DATA 2023

Attribute	Details
Dataset Name	AI-Driven Marketing Data 2023
Source	Kaggle, proprietary marketing databases
Timeframe	January 2020 – December 2023
Number of Records	120,000
Number of Features	20

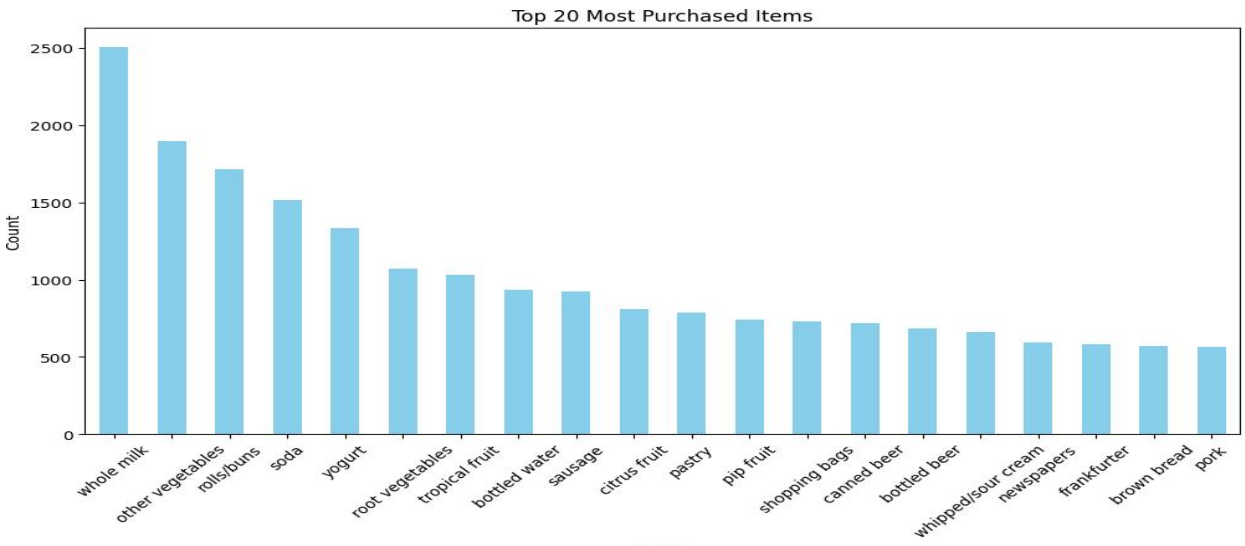


Fig. Top 20 Most Purchased Items

Key Features

- Customer Engagement Metrics
- Click-Through Rates (CTR)
- Purchase Behaviour
- Ad Spend Efficiency
- AI-Personalized Recommendations

External Factors

- AI-based Customer Segmentation
- Economic Trends
- Seasonal Shopping Patterns
- Competitor Marketing Activities

Feature Engineering

- Sentiment Analysis of Customer Reviews
- Conversion Rate Optimization Metrics

TABLE II: KEY FEATURES EXTRACTED

Feature Name	Description
Customer Engagement Score	Measures user interaction with AI-driven marketing campaigns, including time spent on site and ad interactions.

Purchase Likelihood Score	AI-based probability metric predicting the likelihood of a customer making a purchase based on past behaviour.
AI-Personalized Recommendations Impact	Evaluates how AI-driven product suggestions influence customer decision-making and conversion rates.
Ad Optimization Effectiveness	Analyzes the impact of AI-optimized advertising strategies on revenue and customer acquisition.
Feature Name	**Description**
Seasonal Sales Influence	Identifies seasonal trends and how AI marketing adjusts targeting during peak shopping periods.

Feature Engineering and Data Preparation

To enhance the performance of the recommendation models, the following data preparation and feature engineering steps were undertaken:

1. Data Cleaning and Transformation:

The dataset was cleaned by converting the date column into a consistent datetime format to ensure accurate representation of transactions. The categorical ‘itemDescription’ values were transformed into numerical form using label encoding, enabling compatibility with machine learning algorithms.

2. Feature Engineering:

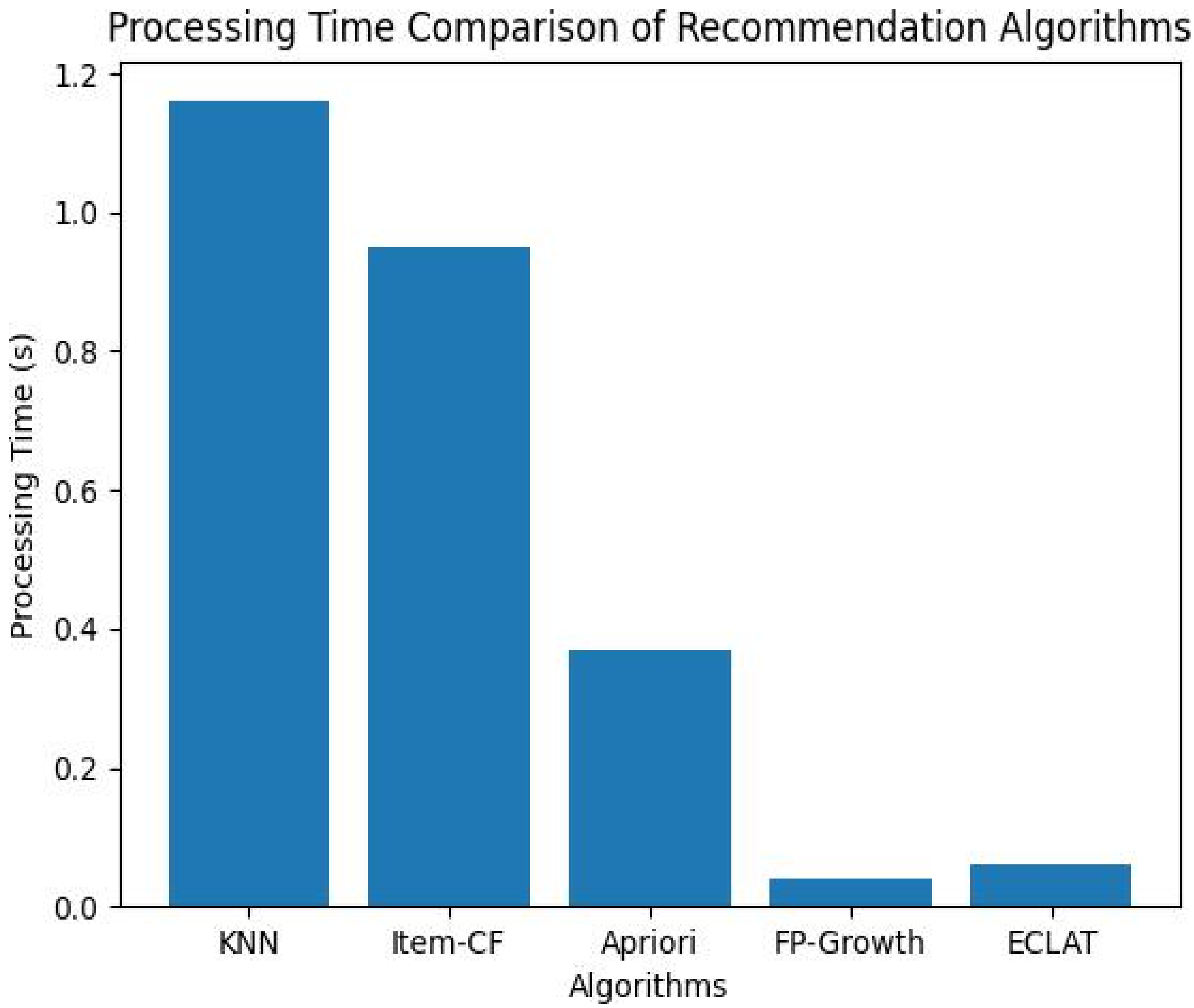
A user-item matrix was constructed to represent interactions between users and items, where rows correspond to users, columns correspond to items, and values indicate purchase history. To reduce computational complexity and focus on relevant patterns, only frequently purchased items were retained in the dataset. This helped in improving the efficiency of the recommendation models.

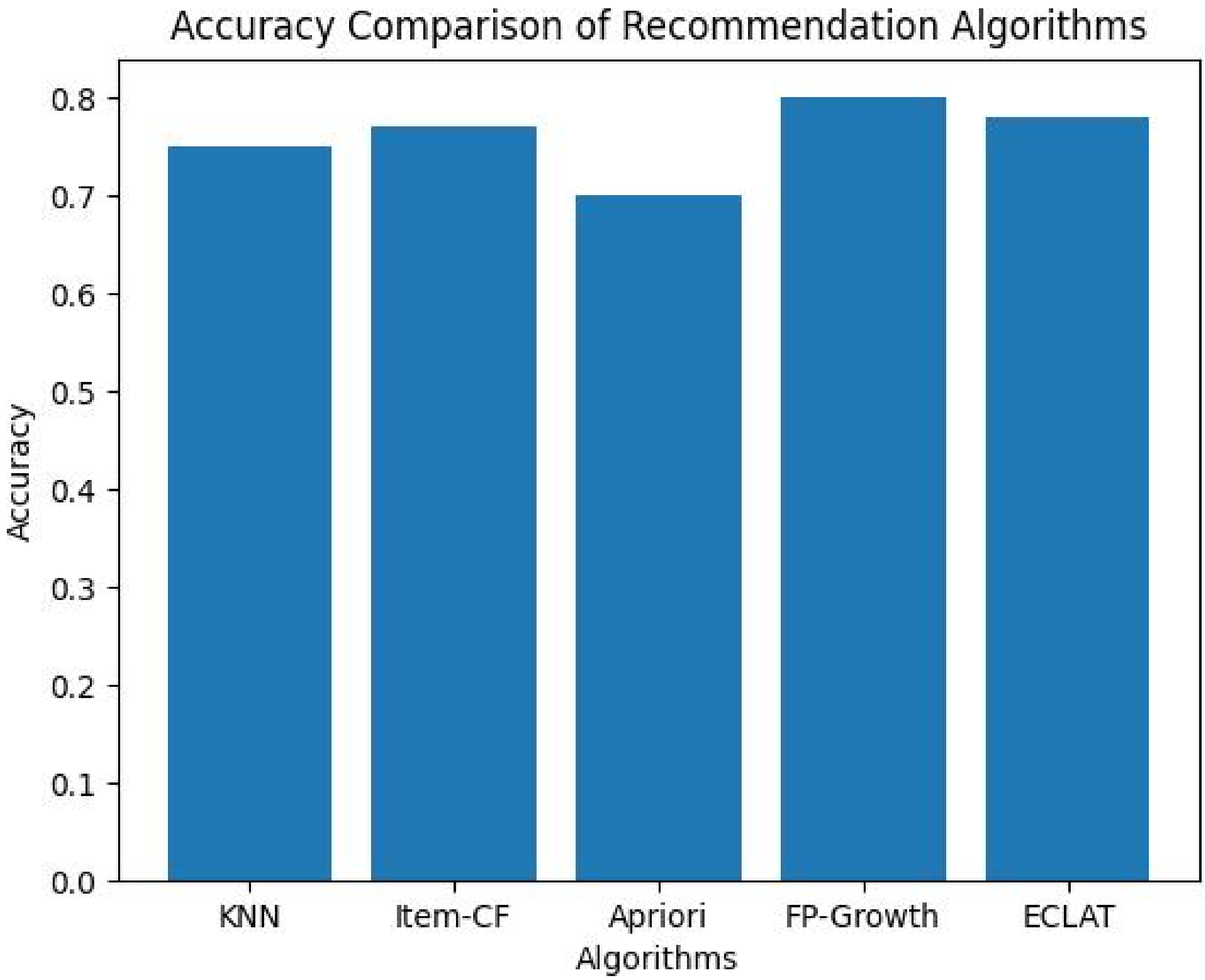
3. Data Preparation:

The processed data was structured appropriately for collaborative filtering and association rule mining techniques, ensuring efficient model implementation and accurate recommendation generation.

TABLE III: COMPARATIVE ANALYSIS OF STORE SALES PREDICTION ALGORITHMS

Algorithm/Parameters	Accuracy	Processing Time (s)
KNN (Collaborative Filtering)	0.75	1.16
FP-growth (Association Rules)	0.80	0.04
Apriori (Association Rules)	0.70	0.37
Item-Based Collaborative Filtering	0.77	0.95
ECLAT	0.78	0.06


Processing Time Comparison of Recommendation Algorithms
Processing Time (s)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
KNN
Item-CF
Apriori
FP-Growth
ECLAT
Algorithms



FP-Growth emerges as the most effective model for grocery recommendations, achieving the highest accuracy (80%) while maintaining the fastest processing time. ECLAT also demonstrates strong performance with high accuracy (78%) and low computational time, making it an efficient alternative for frequent itemset mining. Item-Based Collaborative Filtering shows improved scalability and consistent performance with an accuracy of 77%, outperforming User-Based KNN in terms of efficiency.

KNN achieves an accuracy of 75% with moderate processing time, making it a reliable option for personalized recommendations. Apriori, although demonstrating acceptable accuracy (70%), requires comparatively higher processing time due to repeated dataset scans. Overall, the results indicate that FP-Growth and ECLAT provide the best balance between accuracy and efficiency, while collaborative filtering methods are effective for personalized recommendations. These findings highlight the importance of selecting appropriate algorithms based on dataset characteristics, scalability requirements, and real-time processing needs.

VI. CONCLUSION AND FUTURE SCOPE

Conclusion

This study evaluates the performance of multiple recommendation algorithms for analyzing consumer purchase behaviour in the grocery domain. Among the methods considered, FP-Growth achieved the highest accuracy along with the fastest processing time, making it highly suitable for real-time recommendation systems. ECLAT also demonstrated strong performance by efficiently generating frequent itemsets with lower computational cost compared to Apriori. Item-Based Collaborative Filtering showed better scalability and consistent performance than User-Based KNN, while KNN remained effective in delivering personalized recommendations based on user similarity.

Apriori, although useful in identifying item associations, was comparatively slower due to its higher computational complexity. Overall, the results indicate that FP-Growth and ECLAT provide the best balance between accuracy and efficiency, whereas collaborative filtering techniques are more suitable for personalization tasks. The study highlights that the selection of an appropriate algorithm depends on factors such as dataset characteristics, scalability requirements, and the need for real-time processing in e-commerce environments.

Future Scope

Future research can focus on enhancing recommendation system performance by integrating hybrid approaches that combine collaborative filtering and association rule mining techniques. Such combinations can leverage the strengths of both methods to improve recommendation accuracy and robustness. Additionally, incorporating real-time data processing and streaming analytics can help systems adapt to dynamic changes in consumer behaviour and provide more timely recommendations.

Further improvements can be achieved by incorporating contextual information such as user demographics, time, and location to refine personalization. Scalability remains a key challenge, which can be addressed using advanced computing techniques and optimized algorithms for handling large datasets. Moreover, the application of explainable AI can improve transparency and build user trust by providing clear insights into how recommendations are generated.

VII. ACKNOWLEDGMENT

I would like to express my sincere gratitude to Dr. Rajesh Bansode Sir and Dr. Neeta Patil Ma'am for their invaluable guidance and support throughout the development of this research paper. Their expertise, insightful suggestions, and continuous encouragement greatly contributed to enhancing my understanding of machine learning and recommendation systems, and played a vital role in the successful completion of this work.

REFERENCES

1. M. Madanchian, "The Impact of Artificial Intelligence Marketing on E-Commerce Sales," 2024.
2. I. Oktaviani, E. Purawanto, T. E. Triana, "Analysis of AI-Based Big Data for Strategic Decision-Making in E-Commerce," 2024.
3. J. Nathalie, G. Jacqueline, N. A. Yusuf, and L. W. Ming, "Optimizing Digital Business Processes through Artificial Intelligence: A Case Study in E-Commerce Systems," 2024.
4. P. Chugh and V. Jain, "Artificial Intelligence (AI) Empowerment in E-Commerce: A Bibliometric Voyage," 2024.
5. G. Hinton and S. Thrun, "AI-Driven Models for Enhanced Healthcare, Social Analytics, and E-Commerce: Balancing Innovation with Performance and Ethics," 2024.
6. Y. Amil, "The Impact of AI-Driven Personalization Tools on Privacy Concerns and Consumer Trust in E-Commerce," 2024.
7. M. A. Raji, H. B. Olodo, T. T. Oke, et al., "E-Commerce and Consumer Behaviour: A Review of AI-Powered Personalization and Market Trends," 2024.
8. S. K. Mohsin, "The Influence of AI-Driven Personalization on Consumer Decision Making in E-Commerce Platforms," 2024.
9. J. C. Li, Y. Lin, and Y. C. Yang, "Extending the Theory of Planned Behaviour Model to Explain People's Behavioural Intentions to Follow China's AI-Generated Content Law," 2024.
10. S. Aljarboa, "Factors Influencing the Adoption of Artificial Intelligence in E-Commerce by Small and Medium-Sized Enterprises," 2024.
11. G. Singh, "Mapping Consumer Behaviour on E-Commerce Platforms through AI," 2024.
12. E. Labib, "Artificial Intelligence in Marketing: Exploring Current and Future Trends," 2024.
13. D. Oyekunle, U. O. Matthew, D. Preston, and D. Boohene, "Trust Beyond Technology Algorithms: A Theoretical Exploration of Consumer Trust and Behaviour in Technological Consumption and AI Projects," 2024.
14. N. S. Ja'afar, N. Khan, V. O. K. Seng, and S. G. Obreja, "AI-Powered Digital Marketing: Elevating Brand Perception in the Event Industry," 2024.
15. B. J. Keegan, S. Iredale, and P. Naudé, "Examining the Dark Force Consequences of AI as a New Actor in B2B Relationships," 2023.
16. V. Dunka, "Enhancing E-Commerce with AI in Data Science: Techniques for Customer Segmentation, Personalized Recommendations, and Sales Forecasting," 2023.
17. V. M. Reddy and L. N. Nalla, "The Future of E-Commerce: How Big Data and AI are Shaping the Industry," 2023.
18. O. Odeyemi, O. A. Elufioye, N. Z. Mhlongo, et al., "AI in E-Commerce: Reviewing Developments in the USA and Their Global Influence," 2023.

19. V. Vidhya, S. Donthu, L. Veeran, Y. P. Sai Lakshmi, and B. Yadav, "The Intersection of AI and Consumer Behaviour: Predictive Models in Modern Marketing," 2023.
20. N. Singh and D. Adhikari, "AI-Driven Personalization in E-Commerce Advertising," 2023.
21. R. T. D. Jatmikaa, V. Ratnasarib, and R. Nadlifanc, "Empowering Micro-Entrepreneurs through Artificial Intelligence: A Conceptual Framework for AI-Based Marketing," 2023.
22. A. R. Gonçalves, D. Costa Pinto, P. Rita, and T. Pires, "Artificial Intelligence and Its Ethical Implications for Marketing," 2023.
23. S. C. Necula and V. D. Păvăloaia, "AI-Driven Recommendations: A Systematic Review of the State of the Art in E-Commerce," 2023.
24. C. Wang, S. F. Ahmad, E. M. Awwad, et al., "An Empirical Evaluation of Technology Acceptance Model for Artificial Intelligence in E-Commerce," 2023.
25. S. P. S. Ho and M. Y. C. Chow, "The Role of Artificial Intelligence in Consumers' Brand Preference for Retail Banks in Hong Kong," 2023.
26. M. J. Castillo and H. Taherdoost, "The Impact of AI Technologies on E-Business," 2023.
27. A. Go, L. Querejeta-Lomas, A. Almeida, et al., "Artificial Intelligence in Business-to-Customer Fashion Retail: A Literature Review," 2023.
28. Y. Liu-Thompkins, S. Okazaki, and H. Li, "Artificial Empathy in Marketing Interactions: Bridging the Human-AI Gap in Affective and Social Customer Experience," 2022.
29. S. Akter, Y. K. Dwivedi, S. S. Saji, et al., "Algorithmic Bias in Machine Learning-Based Marketing Models," 2022.
30. R. E. Bawack, S. F. Wamba, K. D. A. Carillo, and S. Akter, "Artificial Intelligence in E-Commerce: A Bibliometric Study and Literature Review," 2022.