Keerti Pal, 2025, 13:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Customer Product Choice Recommendation by Association Rules and Learning Models

Keerti Pal[1], Jayshree Boaddh[2], Rahul Patidar[3]
[1]MTech Scholar, Vaishnavi Institutes of Technology and Science, Bhopal
[2]HOD, CSE, Vaishnavi Institutes of Technology and Science, Bhopal
[3]Asst. Professor, CSE, Vaishnavi Institutes of Technology and Science, Bhopal

Abstract- Online stores and apps attracts customer at various levels. So approaches need to be more effective by analyzing the behavior of visiting customer. Many of researcher has proposed different models of customer product recommendation system. This paper introduces a novel Customer Product Recommendation by Rules and Ensemble Model (CPRRESM) framework designed to enhance purchase prediction accuracy for small-scale retail stores with limited data resources. The proposed approach integrates Apriori-based association rule mining for pattern discovery, Z-score normalization for feature standardization, and a Gradient Boosting ensemble model for efficient learning. By combining rule-based insights with ensemble learning, CPRRESM effectively captures customer purchasing behavior and dynamic preferences.

Keywords: Shopping Store, Product Recommendation, Feature Optimization.

I. INTRODUCTION

In the early 2000s, data analytics became a pivotal tool for identifying optimal branch locations and service points, driving the banking and financial sectors to develop more personalized and digital services [1]. As digitalization advanced and customer behavior evolved, the industry recognized the need to adapt to these transformations. Today, with the vast number of transactions processed daily by banks, traditional data processing methods are no longer sufficient to manage the complexity of analytical tasks and transaction workflows. In modern business environments, effectively leveraging data insights has emerged as a critical factor for banks to maintain a competitive edge.

The growing importance of data analytics stems from the fact that data act as the fundamental raw material for generating information, which directly influences decision-making. To extract economic value from data, it must first be refined and analyzed using sophisticated analytical approaches [2]. Consequently, machine learning (ML) has become a vital technology, enabling the cleaning, processing, analysis, and discovery of hidden patterns within datasets. Like many other industries that capitalize on data-driven strategies, the banking sector has increasingly integrated ML techniques into its business operations.

Recommender systems refer to software tools and methodologies that utilize user preferences and historical data to suggest relevant products or items. These systems support users in making informed choices by presenting personalized recommendations [3]. They rely on various forms of actively collected data to generate and refine these suggestions, with the nature of the data depending on the type of recommender system employed. Moreover, recommender systems have gained prominence through their ability to exploit the vast amount of information available on Online Social Networks (OSNs), offering users meaningful recommendations on products, services, and other entities [4].

Rest of paper was brief in four sections. Second section brief the related work done by researchers in previous years. Further porposed model was detailed with block diagram. Futher proposed model is cmapred with existing work under same dataset. Finally whole work was concluded.

II. RELATED WORK

Lifeng Kang et al. [5] introduced a fusion recommendation algorithm that leverages frequent item set mining to address key challenges in recommendation systems. The proposed approach focuses on compressing large commodity datasets and identifying frequent item sets to enhance recommendation efficiency. The paper first highlights the limitations of existing fusion recommendation methods based on frequent item set mining, including redundant rule generation, low accuracy, and limited exploration of deeper user–product relationships.

To overcome these issues, the authors propose an improved algorithm that filters commodity datasets, computes user–item interest rankings, and formulates recommendation rules for similar products. Experimental results demonstrate that the algorithm effectively adapts to users' evolving preferences and captures their dynamic interests in real time.

Ma et al. [6] explored the determinants of consumer engagement in live shopping environments. Their study employs structural equation modeling (SEM) to examine the factors influencing the non-transactional dimensions of consumer engagement in live-stream commerce. Additionally, fuzzy-set qualitative comparative analysis (fsQCA) is used to identify combinations of factors that enhance engagement levels. The findings indicate that visual cues, personal involvement, and streamer identification significantly contribute to users' emotional and informational support during live-streaming sessions.

Xu et al. [7] proposed a model for understanding cultural heritage product purchases, drawing on the ABC theory of attitudes to explore how intangible cultural heritage items affect consumers' purchase intentions in live-streaming contexts. The study applies confirmatory factor analysis (CFA) and structural equation modeling (SEM) for data analysis and hypothesis validation. Results show that consumer satisfaction plays a mediating role in linking product quality to purchase intention, emphasizing the importance of product authenticity and perceived value.

Kumar et al. [8] examined the application of the Markov chain model for predicting hidden states through transition probabilities in e-commerce systems. The model assists in identifying usability patterns and issues via behavioral metrics and user interaction analysis. The study highlights the importance of machine learning and association rule mining within this framework. Furthermore, the authors emphasize that data analytics plays a pivotal role in e-commerce for inventory optimization, fraud detection, and customer personalization, leveraging historical and statistical data to gain a competitive advantage.

Shi et al. [9] introduced a decision-making framework that integrates a Fuzzy Decision Support System (FDSS) with a Fuzzy Analytic Hierarchy Process (FAHP). This framework is designed to prioritize design elements, develop cultural and creative design components, and identify key criteria that influence user requirements. The study demonstrates that the proposed approach assists industrial designers in developing more effective color schemes for cultural and creative products by incorporating collective user visual preferences into purchase intention through multiuser decision consistency.

Dogan et al. [10] enhanced the traditional Association Rule Mining (ARM) methodology by incorporating fuzzy set theory, resulting in a Fuzzy Association Rule Mining (FARM) approach. Unlike conventional ARM, which considers only item occurrences, the improved method also accounts for sales quantities. By employing the Apriori algorithm, FARM can capture customer preferences from historical transaction data, allowing e-commerce platforms to identify fuzzy association rules and recommend similar products that align with customers' purchasing needs and quantities.

Cach N. Dang et al. [11] developed and assessed a hybrid recommendation system that combines Collaborative Filtering (CF) with Sentiment Analysis (SA) for a more comprehensive recommendation process. The system employs a flexible architecture that integrates advanced Deep Learning (DL) models for sentiment analysis and utilizes improved feature extraction techniques. Experimental evaluations using two benchmark datasets demonstrate that integrating CF with sentiment-driven DL models significantly enhances recommendation accuracy and overall system performance. However, the study notes that this approach is computationally demanding and requires substantial processing resources.

Ching Hsu and An-Hung Liao [12] proposed a hybrid recommendation framework that merges Sentiment Analysis (SA) and Machine Learning (ML) within a chatbot-based system. The framework includes multiple functional modules, and its implementation was tested using a sentiment-based article recommendation Linebot.

This Linebot uses an API interface and a webhook mechanism, enabling smooth system activation and interaction. The study compared the performance of four ML algorithms and two DL models within a Spark cloud computing environment. Experimental results revealed that the decision tree algorithm outperformed others in terms of accuracy and processing efficiency for sentiment analysis tasks. However, the research's main limitation lies in its lack of integration of decision-making systems that incorporate human behavioral or psychological factors.

III. PROPOSED WORK

This section brief proposed Customer Product Recommendation by Rules and Ensemble Model (CPRRESM). User purchase prediction depends on multiple factors, and improving the predictive capability for small shopping stores is crucial. These stores often suffer from limited resources and datasets, which reduces research interest and restricts advanced analysis. Therefore, accurate customer purchase prediction becomes essential. Figure 1 illustrates the overall workflow of CPRRESM, followed by a detailed explanation of each processing stage.

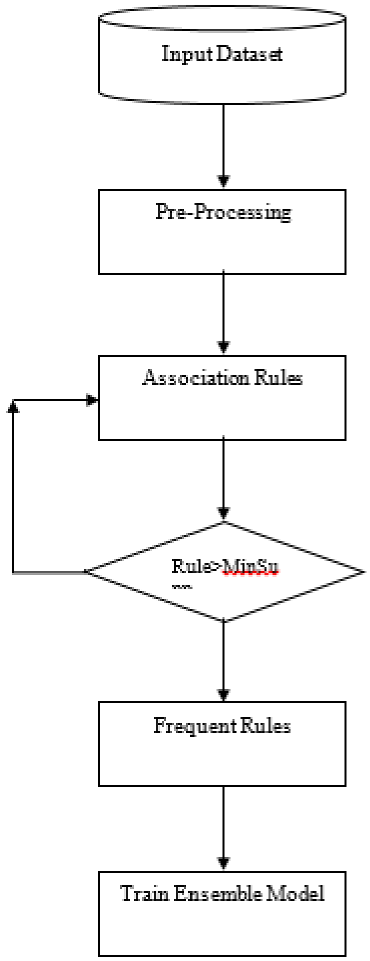

Fig. 1 Proposed CPRRESM model flow diagram.

Input Shopping Store Data

The dataset used in this study consists of transactional information from a retail shopping store. Each record includes essential customer and transaction details such as gender, date, amount spent, purchase type, and location. The dataset contains approximately 99,000 entries with 10 attributes per transaction [13]. Before model training, data preprocessing is applied to ensure quality and relevance.

Pre-Processing of Dataset

Each transaction in the dataset includes a unique row identifier and customer ID, which are irrelevant for prediction and thus removed [14]. Similarly, the purchase date attribute was excluded, as it does not directly aid in discovering meaningful behavioral patterns.

Categorical data such as purchase type or product category were converted into numerical values using label encoding—for example, "Clothing" is encoded as 1, "Footwear" as 2, and so on. The preprocessed dataset is represented as:

PIRSD・PreProcess(IRSD)
where ISSD denotes the Input Shopping Store Data and PISSD represents the Preprocessed Input.

Pattern Extraction Using Apriori Algorithm
Patterns within the transactional data reveal significant relationships among attributes that directly influence purchasing behavior [15]. To uncover these patterns, the Apriori algorithm is used to generate frequent itemsets and association rules.

The Apriori process involves:

- **Frequent 1-Itemset Generation:**
  Identify individual items that satisfy the minimum support threshold.
- **Candidate Itemset Generation:**
  Combine frequent (k–1)-itemsets to form candidate k-itemsets.
- **Support Calculation and Pruning:**
  Compute support for each candidate itemset and remove those that do not meet the support requirement.
- **Rule Formation:**
  Generate association rules from the frequent itemsets that meet a predefined confidence threshold [16].

This approach enables the discovery of co-occurrence patterns such as "Customers who buy shoes often buy socks," which can enhance predictive modeling.

Normalization Using Z-Score Standardization
To ensure that all selected features contribute equally to the learning process, the dataset is normalized using Z-score normalization [17]. This method centers each feature around zero with unit variance, maintaining the distribution shape while mitigating scale bias.

The normalization is computed as:

$$FFs = \frac{X - Mean}{Standard\_Deviation}$$

Learning Using Gradient Boosting Ensemble Model
The processed and normalized dataset is used to train an ensemble learning model—specifically, the Gradient Boosting algorithm [18]. Gradient Boosting is a powerful ensemble approach that builds multiple weak learners (typically decision trees) sequentially, where each new tree focuses on correcting the errors made by previous ones.

Working Mechanism:

- Initialize the model with a base prediction (mean outcome for regression or log-odds for classification).
- Compute residuals (errors) based on current predictions.
- Train a new decision tree on these residuals.
- Add the new tree's weighted contribution to the ensemble model.
- Repeat the process until the stopping criterion (e.g., minimum error or maximum number of trees) is met.

Each iteration aims to minimize a differentiable loss function, enhancing both bias reduction and generalization performance.

Feature Sampling:
To improve diversity, Gradient Boosting can also incorporate random feature sampling—different subsets of attributes are used for different models.

1. Proposed Algorithm
2. PIRSD ← PreProcess(IRSD)
3. FFP ← Apriori_Frequent_Pattern(PISSD) // FFP: Frequent Feature Patterns
4. For each pattern P in FFP
5. If Support(P) ≥ MinSup
6. Frequent_P ← P
7. EndIf
8. EndLoop
9. 7. NDS ← ZScore_Normalization(Frequent_P) // NDS: Normalized Dataset
10. [TrainSet, TestSet] ← Split_Data(NDS)
11. GB ← Initialize_GB(Tree_Count, Learning_Rate)
12. Train_GB ← Train_GB(GB, TrainSet)
13. Predict ← Test_GB(Train_GB, TestSet)
    Evaluate(Predict)

The proposed UPP-EL model integrates pattern discovery through the Apriori algorithm, Z-score normalization, and Gradient Boosting ensemble learning for robust customer purchase prediction. This combination enhances interpretability, balances feature influence, and achieves high prediction accuracy for small-scale shopping store datasets.

IV. EXPERIMENT AND RESULT

Experimental Setup

This section provides a detailed evaluation of the experimental outcomes achieved using the proposed approach. All algorithms and performance assessments were implemented and executed within the MATLAB environment. The experiments were carried out on a computing system equipped with an Intel Core i3 processor (2.27 GHz), 4 GB RAM, and running Windows 7 Professional.

The experimental dataset, comprising 99,000 shopping transactions, was sourced from [19]. The performance of the proposed method was compared with the existing model titled "Exploratory Data Analysis Machine Learning Prediction (EDAMLP)" as referenced in [20].

Evaluation Parameter

To test outcomes of the work following are the evaluation parameter such as Precision, Recall and F-score.

Precision = TP / (TP+ FP)

Recall = TP / (TP + TN)

F-score = 2 * Precision * Recall / (Precision + Recall)

Results

Table 1 Customer Product Recommendation models Precision values.

Dataset Size	EDAMLP	CPRRESM
600	0.6293	0.8577
1200	0.6395	0.8565
1800	0.6284	0.8503
2400	0.6434	0.8437
3000	0.6453	0.8417
4000	0.6417	0.8463

Table 1 shows precision value of shopping category prediction against different experimental dataset percentage. It was found that use of ensemble learning model has increases the precision value by 24.89% as compared to EDAMLP.

Table 2 Customer Product Recommendation models Recall values.

Dataset Size	EDAMLP	CPRRESM
600	0.3559	0.7946
1200	0.3167	0.8055
1800	0.314	0.8064
2400	0.3212	0.8033
3000	0.317	0.8036
4000	0.3148	0.8022

Recall values of shopping category prediction models were shown in table 2, it was found that use of Z-Score normalization increases the value. Fig. 2 shows Use of ensemble learning learning of customer behavior has increases high impact on prediction output.

Table 3 Customer Product Recommendation models F-Measure values.

Dataset Size	EDAMLP	CPRRESM
600	0.4547	0.8249
1200	0.4236	0.8302
1800	0.4187	0.8278
2400	0.4285	0.823
3000	0.4251	0.8222
4000	0.4224	0.8237

Table 3 shows different experimental data percentage f-measure values for customer product purchase category prediction. It was found that, use of association rule for getting the patterns has improved the work perforance by removing the unnecessary information.

Table 4 Customer Product Recommendation models Accuracy values.

Dataset Size	EDAMLP	CPRRESM
600	50.67	71
1200	46.33	71.92
1800	46.17	71.5
2400	46.42	70.87
3000	45.73	70.8
4000	45.4	70.95

Table 4 shows Prediction accuracy of customer product category recommendation models. It was found that use of association rules for pattern extraction has enhanced the leaning of ensemble model. Fig. 4 shows that Fuzzy normalization of input learning model has increases the prediction accuracy by 34.26% as compared to EDAMLP.

V. CONCLUSION

The proposed CPRRESM model successfully integrates data preprocessing, frequent pattern extraction, and ensemble-based learning to enhance customer purchase prediction in small shopping store datasets. The use of association rules effectively identifies meaningful item relationships, while Z-score normalization ensures balanced feature contribution during learning. Experimental results confirm that the ensemble approach significantly improves prediction accuracy and robustness over conventional models such as EDAMLP. Thus, CPRRESM provides an efficient, interpretable, and scalable solution for personalized product recommendation and can serve as a valuable tool for data-driven decision-making in small retail environments. Experimental evaluation conducted in the MATLAB environment demonstrates that CPRRESM achieves superior precision, recall, F-measure, and accuracy compared to the existing EDAMLP model, improving precision by approximately 24.89% and overall prediction accuracy by 34.26%.

REFERENCES

1. I G. Cossatin, N. Mauro, G. Izzi and L. Ardissono, "Synchronized multi-list user interfaces for fashion catalogs", Proc. 31st ACM Conf. User Model. Adaptation Personalization, pp. 224-228, Jun. 2023.
2. D. Starke, E. Asotic, C. Trattner and E. J. van Loo, "Examining the user evaluation of multi-list recommender interfaces in the context of healthy recipe choices", ACM Trans. Recommender Syst., vol. 1, no. 4, pp. 1-31, Dec. 2023.
3. El Majjodi, A. D. Starke and C. Trattner, "Nudging towards health? Examining the merits of nutrition labels and personalization in a recipe recommender system", Proc. 30th ACM Conf. User Model. Adaptation Personalization, pp. 48-56, Jul. 2022.
4. Andaur, J.M.R.; Ruz, G.A.; Goycoolea, M. Predicting Out-of-Stock Using Machine Learning: An Application in a Retail Packaged Foods Manufacturing Company. Electronics 2021, 10, 2787.
5. Lifeng Kang, Yankun Wang, Efficient and accurate personalized product recommendations through frequent item set mining fusion algorithm, Heliyon, Volume 10, Issue 3, 2024.
6. Ma, Y., A socio-technical analysis of factors affecting consumer engagement in livestream shopping: Evidence from structural equation modeling and fuzzy set qualitative comparative analysis. Telematics Inform. 87, 102094 2024.
7. Xu, Y. Q., Hasan, N. A. M., & Jalis, F. M. M. (2024). Purchase intentions for cultural heritage products in E-commerce live streaming: An ABC attitude theory analysis. Heliyon, 10(5).
8. Kumar, B.; Roy, S.; Sinha, A.; Iwendi, C.; Strážovská, Ľ. E-commerce website usability analysis using the association rule mining and machine learning algorithm. Mathematics 2022, 11, 25.
9. Shi, L., Yang, X. Personalized Recommendation Algorithm for Cultural and Creative Products Based on Fuzzy Decision Support System. Int J Comput Intell Syst 18, 116 (2025).
10. Dogan, O., Kem, F.C. & Oztaysi, B. Fuzzy association rule mining approach to identify e-commerce product association considering sales amount. Complex Intell. Syst. 8, 1551–1560 (2022).
11. Dang CN, Moreno-García MN, Prieta FDL. An approach to integrating sentiment analysis into recommender systems. Sensors. 2021;21(16):5666.
12. Hsu IC, Liao AH. (2022). Sentiment-based chatbot using machine learning for recommendation system.
13. K. Dhanushkodi, A. Bala, N. Kodipyaka and V. Shreyas, "Customer Behavior Analysis and Predictive Modeling in Supermarket Retail: A Comprehensive Data Mining Approach," in IEEE Access, vol. 13, pp. 2945-2957, 2025.

14. M. Copik, T. Grosser, T. Hoefler, P. Bientinesi and B. Berkels, "Work-Stealing Prefix Scan: Addressing Load Imbalance in Large-Scale Image Registration," in IEEE Transactions on Parallel and Distributed Systems, vol. 33, no. 3, pp. 523-535, 1 March 2022.
15. Azeez, N.A.; Ayemobola, T.J.; Misra, S.; Maskeliūnas, R.; Damaševičius, R. Network Intrusion Detection with a Hashing Based Apriori Algorithm Using Hadoop MapReduce. Computers 2019, 8, 86.
16. Eni Heni Hermaliani et al 2020 J. Phys.: Conf. Ser. 1641 012070
    N. Fei, Y. Gao, Z. Lu and T. Xiang, "Z-Score Normalization, Hubness, and Few-Shot Learning," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 142-151,
17. Phanbua, P.; Arwatchananukul, S.; Hristov, G.; Temdee, P. CTGAN-Augmented Ensemble Learning Models for Classifying Dementia and Heart Failure. Inventions 2025, 10, 101.
    https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset
18. K. Dhanushkodi, A. Bala, N. Kodipyaka and V. Shreyas, "Customer Behavior Analysis and Predictive Modeling in Supermarket Retail: A Comprehensive Data Mining Approach," in IEEE Access, vol. 13, pp. 2945-2957, 2025