International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Shopping Store Product Recommendation by Patterns and Random Forest Model

Sonu Kumar[1], Yogesh Rai[2]

[1]M.Tech. Scholar Technocrats Institute of Technology Excellence
[2]Asst. Professor, CSE, Technocrats Institute of Technology Excellence

Abstract
Shopping stores need to grow that directly depends on the customer pattern. Many of researchers have done lot of work and it was found that doing recommendation system need improvement for small store customers. This paper has proposed a model that performs preprocessing steps to remove unwanted information. Further paper has extract patterns from the processed data. Low frequent patterns were removed and highly frequent patterns were used for the learning. Random forest model was used for the learning of prefixscan. Experiment was done on real dataset and results shows that proposed User Purchase Prediction by Patterns and Random Forest (UPPPRF) model has improved the prediction accuracy by % as compared to existing models.
Keywords:-

Keywords: Shopping Store, Product Recommendation, Feature Optimization**.**

1. Introduction

With the explosive growth of the internet, businesses are increasingly compelled to build a robust digital presence to maintain relevance and outperform competitors. The internet has transformed how companies function across both local and global markets, becoming a vital component of modern business operations [1]. This surge in connectivity has driven a fundamental shift in business strategies, particularly in favor of digital transformation and online commercial activities.

One of the most significant changes brought by the digital era is the emergence and rapid growth of e-commerce. Online platforms provide businesses with the ability to market, display, and sell products and services to a broad and diverse audience [2]. Unlike traditional physical stores, e-commerce platforms allow for seamless transactions across different regions, offering consumers unparalleled convenience and access.

To enhance user engagement and drive sales, e-commerce platforms increasingly depend on recommender systems—intelligent tools that suggest relevant products based on user interactions, preferences, or demographics [3]. These systems employ various methods, such as:

- Top-selling product suggestions based on site-wide sales data.
- Demographic-based recommendations customized using user attributes like age or location.
- Behavior-based recommendations derived from a customer's browsing and purchase history.

Such techniques contribute significantly to personalized shopping experiences, helping platforms present tailored content that aligns with individual user interests [4]. This personalization fosters higher satisfaction, repeat visits, and increased conversion rates.

The benefits of recommender systems span three main areas: converting passive users into buyers, encouraging additional purchases through cross-selling, and building long-term customer loyalty by enhancing user experience [5].
However, traditional recommender systems—primarily reliant on historical ratings—face challenges in capturing user preferences accurately. To overcome this, the current research proposes a hybrid model that combines rating data with user trust relationships. By analyzing latent features of users and items and incorporating social trust metrics, the model aims to deliver more precise and context-aware product suggestions, ultimately improving recommendation quality and system reliability.

Related Work

Andaur, J.M.R. et al. [6] addressed the persistent issue of out-of-stock (OOS) events from a manufacturing perspective by conducting a focused case study within a packaged foods manufacturing company located in Latin America. Their work aimed to develop machine learning-based systems that could detect and predict OOS occurrences more efficiently and with higher accuracy. Two distinct models were developed and evaluated: the first utilized a single Random Forest classifier trained on a balanced dataset, while the second adopted an ensemble approach combining six diverse classification algorithms. The data used to train and validate these models were sourced from both the manufacturer's internal transaction system and physical audit records, ensuring a robust and realistic representation of the operational environment. A key innovation of this study lies in the inclusion of newly introduced predictor variables, which enhanced the model's performance in identifying OOS patterns. These improvements allowed for successful deployment and testing of the predictive system in a real-world industrial setting, resulting in improved forecasting accuracy and operational responsiveness.
In another significant contribution, MicolPolicarpo et al. [7] conducted a comprehensive literature review examining the use of machine learning techniques across various e-commerce functions. These included predicting purchases, repurchases, and the likelihood of product returns. Their principal contribution was the introduction of a novel taxonomy classifying the diverse goals addressed by ML in e-commerce. However, the study notably omitted a detailed exploration of return prediction mechanisms within this framework, pointing to a potential gap for future research.
N. Su et al. [8] proposed a weakly-labeled, regularized topic model designed to capture intricate patterns in client purchasing behavior. This approach begins with extracting a wide range of features—textual, semantic, and visual—that encapsulate a customer's individual preferences. These are fed into a probabilistic topic modeling framework where each user's interest is modeled as a distribution over product categories. To address data sparsity, especially in cases with limited user histories, a regularization component was added. The model also employs Kullback–Leibler (KL) divergence to calculate preference similarity between users, constructing an affinity graph and identifying purchasing clusters through dense graph mining.
The study presented in [9] explored the deployment of both batch and real-time recommendation engines using the RP3Beta model, a lightweight yet effective graph-based collaborative filtering method. Although the model is relatively simple, it consistently outperforms more complex algorithms. Its design is adaptable, enabling integration into any recommendation system that supports item-to-item relationships, thus making it particularly useful in environments where user-specific data may be limited.
In [10], the authors built a knowledge graph using instructional resources to enhance a recommendation

system in an educational setting. By combining explicit user-resource interactions with auxiliary semantic connections in the knowledge graph, and applying an attention mechanism to capture individualized learner preferences, the system generates enriched user and resource representations. These are then used to compute the inner product for preference prediction.

Lastly, [11] introduced a hybrid AI-driven business intelligence system for predicting food sales in local markets. By integrating neural networks, historical sales data, and individual client profiles, the model delivers low mean squared error and variance, offering grocery retailers precise insights into customer preferences and future demand.

K. Dhanushkodiet. al. in [12], proposed a predictive analysis techniques with Support Vector Machine (SVM) and Neural Network classifiers to accurately predict customer segments. Both models exhibited impressive performance metrics, highlighting their effectiveness in classifying instances into the correct segments. Additionally, sequential pattern mining using the PrefixSpan algorithm revealed frequent sequences of customer purchase behaviour, offering valuable insights for targeted marketing and personalized recommendations.

2. Proposed Work:

User purchase prediction depends on different factors out of those small shopping stores need to be improved a lot. Limited resource and data attract less researchers hence prediction of customer purchase is important. This section brief the proposed User Purchase Prediction by Patterns and Random Forest (UPPPRF). Fig. 1 shows the block diagram of various steps involved in UPPPRF and explanation of each block was done as well.

Input Shopping Store

This work has uses transaction information of the shopping store having basic information of customer with fruitful information like Gender, Date, amount of purchase, type of purchasing, place, etc. Input dataset have 99000 number of entries with 10 attributes of each transaction. Input dataset is pre-process further to remove unwanted information for ther learning.

Pre-Processing of Dataset

Input shopping store dataset have unique id number for each row that need to be removed as this doesn't make any use for the learning or prediction. Further each customer also have id that is also an unwanted information. In this data date of purchase was removed as this doesn't support in finding the patterns [10, 11]. As input data have have some text so a numeric value is assign to each unique text present in a column, lets say if user purchase cloth then word 'cloth' is assigned identify by 1. In similar fashion 'Shoe' assigned by 2.

PISSD←PreProcess(ISSD)

In above eq. ISSD is input Shopping Store and PISSD is preprocessed Input Shopping Store.

Extract Pattern

Set of attributes make pattern in the dataset that directly support in prediction. Paper has generate Pre-fix Scan Association rules from the dataset [13].

Product Rating Dataset

In this dataset product rating feature is present. This can be understand as user U1 has either use or have knowledge or its review for any product id P1then rate it on the basis of his experience. PreFixScan steps are [14].

Step 1: Generate Frequent 1-Itemsets

- Scan the entire dataset to determine the frequency (support count) of each item.
- Discard items that do not meet the minimum support threshold.
- The remaining items form the frequent 1-itemsets (L1).

Step 2: Generate Prefix Tree (Tree)

- Build a prefix tree structure using frequent 1-itemsets:
- Each node represents an item.
- Paths from root to nodes represent item prefixes.
- Each path keeps a record of support counts of items in that path.

Step 3: Recursive Pattern Growth

- Start with each frequent item (prefix).
- Recursively extend the prefix by adding subsequent frequent items that co-occur in the same transactions.

For each prefix:

- Project the database onto the prefix (i.e., extract suffixes of transactions containing the prefix).
- Mine the projected database recursively to find longer frequent itemsets.

Step 4: Generate Frequent Itemsets

- Collect all the frequent itemsets discovered from the prefix tree.
- These itemsets must satisfy the minimum support threshold.

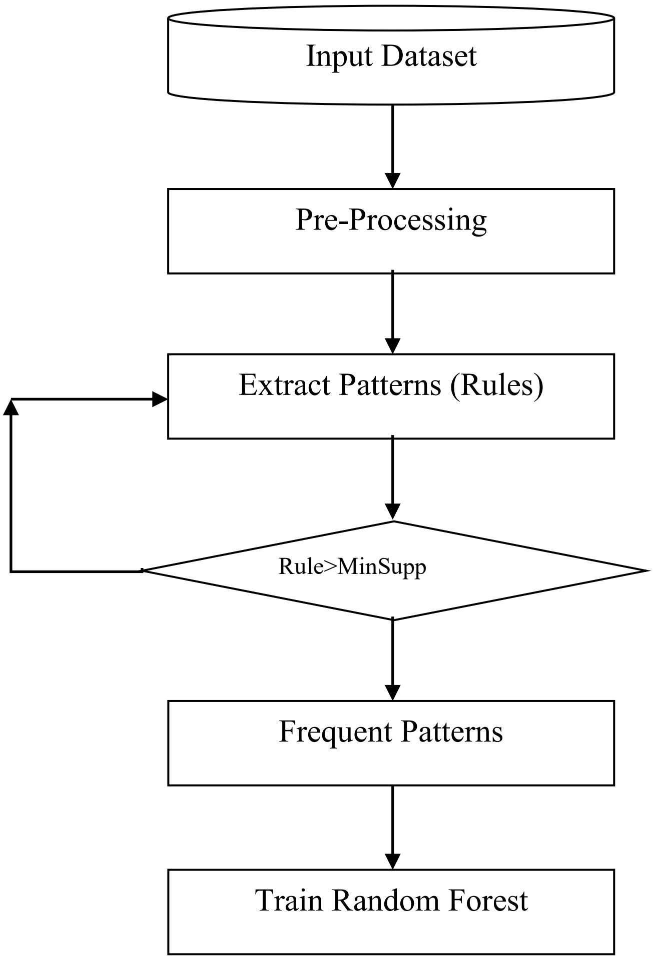

Fig. 1 Proposed model for customer purchase prediction.

Normalization by Fuzzy Min-max

In order to make a equal impact of all selected features normalization of the processed dataset is required. Selected features were normalized by the Min-Max Normalization. This normalization range value between 0 to 1.

Min ←(FFs_{1-n})

Max ←(FFs_{1-n})

$$FFs = \frac{FFs - Min}{Max - Min}$$

Random Forest on Shopping Dataset

The Random Forest algorithm is a robust ensemble method composed of multiple decision trees used for classification and regression tasks [15]. In the context of the shopping dataset. Random Forest is applied to classify customer behaviors, such as purchasing patterns or customer segmentation. The algorithm constructs each tree using a random subset of data (bootstrapping) and random selection of features to ensure diversity among trees. Once all trees are built, a new instance (e.g., a customer's shopping session) is passed through each tree, and the majority vote from all trees determines the final predicted class (e.g., product purchase or no purchase).

Random Feature Selection In the shopping dataset, the Random Forest algorithm selects a unique feature set for each decision tree to enhance diversity and prevent overfitting. For example, one tree might use features like [Customer_Age, Product_Category, Session_Duration], while another might use [Browser_Type, Cart_Items, Payment_Method]. This variability helps improve the model's generalization ability.

CART Algorithm for Tree Construction The Classification and Regression Tree (CART) method is utilized to construct each individual decision tree in the Random Forest [16]. CART can handle both categorical (e.g., product category) and numerical data (e.g., time spent per session). Each tree starts at the root node, recursively splitting into child nodes using binary partitioning. The best split is determined by minimizing an impurity function, typically the Gini Index, which quantifies the homogeneity of the classes.

The Gini Index is calculated as:

GI = ∑(C ∈ {Y,N}) [-P(c,L) log(P_(c,L)) + -P_(c,R) log(P_(c,R))]

Where P_(c,L) and P_(c,R) represent the proportion of transactions classified to the left or right nodes based on a given split.

PruningLarge trees may overfit training data, decreasing accuracy on unseen shopping sessions. Pruning removes unimportant branches from the tree. Using a cost-complexity metric that combines tree size and misclassification rate, the algorithm determines an optimal tree size that balances complexity and prediction accuracy.

Proposed UPPPRF Algorithm

Input: ISSD

Output: Train_RF

PISSD←PreProcess(ISSD)

1. FFP←Frequent_First_Item (PISSD) // FFP:Frequent First Pattern
2. PT←PreFix_Tree(FFP)

3. RPG←Recursive_Pattern(PT)
4. For each pattern P
5. If P>MinSupp
6. Frequent_P←P
7. Endif
8. EndLoop
9. RF←Intial_RF(Tree_Count)
10. Train_RF←Train_RF(RF, Frequent_P)

3. Experiment and Result

Experimental Setup

This section presents a comprehensive analysis of the experimental results obtained for the proposed method. The implementation of all algorithms and performance metrics was done within the MATLAB software framework. The experiments were performed on a computer system featuring an Intel Core i3 processor operating at 2.27 GHz, with 4 GB of RAM, and running the Windows 7 Professional operating system. This work compared with existing model "Exploratory Data Analysis Machine Learning Prediction" EDAMLP in [12].Experimental dataset taken from [17], having 99000 shopping sessions.

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 Precision value of product recommendation system.

Dataset Size	EDAMLP	UPPPRF
500	0.6701	0.8596
1000	0.736	0.862
2000	0.8045	0.842
5000	0.8453	0.8493
8000	0.8919	0.8475
10000	0.8352	0.8458

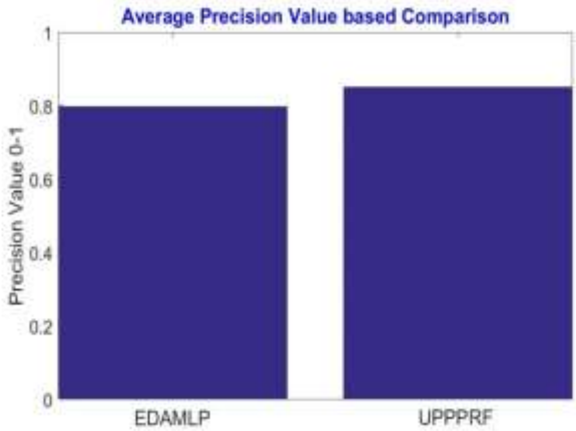

Fig. 1 Average precision value based comparison of product recommendation models.

Table 1 shows precision value of shopping category prediction against different experimental dataset percentage. It was found from fig. 1, that use of random forest model has increases the precision value by 6.32% as compared to EDAMLP.

Table 2 Recall value of product recommendation system.

Dataset Size	EDAMLP	UPPPRF
500	0.8442	0.794
1000	0.7756	0.7988
2000	0.7108	0.8005
5000	0.6625	0.8043
8000	0.6489	0.7972
10000	0.6347	0.7956

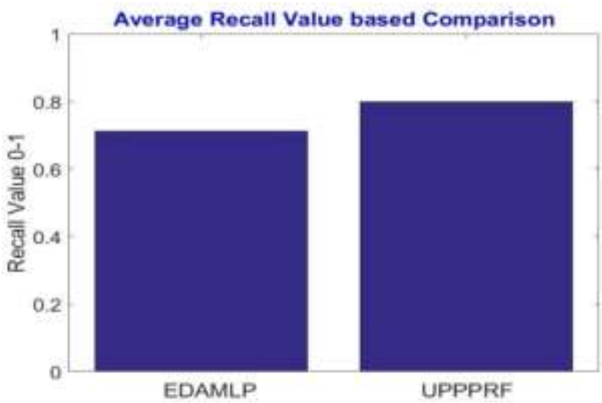

Fig. 1 Average recallvalue based comparison of product recommendation models.

Recall values of shopping category prediction models were shown in table 2, it was found that use of fuzzy min-max normalization increases the value.Fig. 2 shows Use of random forest for the learning of customer behavior has increases high impact on prediction output. Average recall values were improved by the 10.72%.

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Table 3F-Measure value of product recommendation system.

Dataset Size	EDAMLP	UPPPRF
500	0.7471	0.8255
1000	0.7553	0.8292
2000	0.7548	0.8207
5000	0.754	0.8262
8000	0.7513	0.8216
10000	0.7428	0.8199

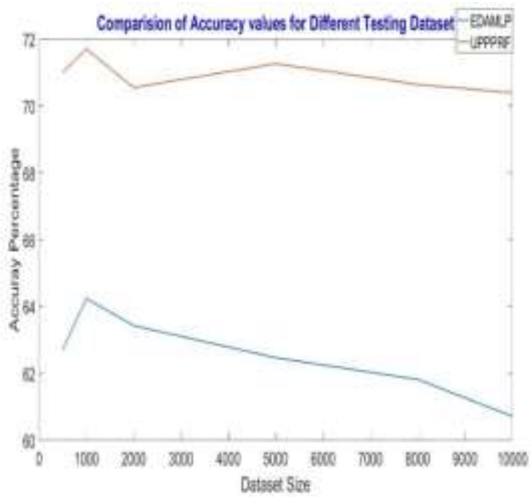

Fig. 3

Comparison of F-Measure for product category prediction. Fig. 4 Average accuracy value of product category prediction

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

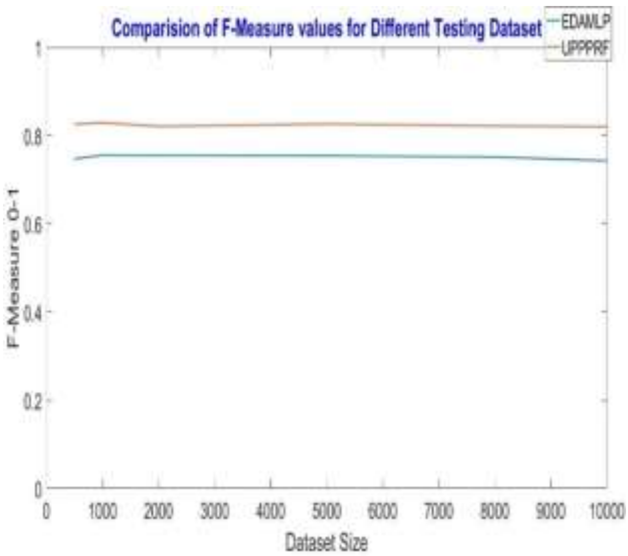

Fig. 4 Average accuracy value of product category prediction.

Table 4Accuracy value of product recommendation system.

Dataset Size	EDAMLP	UPPPRF
500	62.71	71
1000	64.24	71.7
2000	63.43	70.55
5000	62.47	71.26
8000	61.82	70.64
10000	60.72	70.39

Table 4 shows Prediction accuracy of customer product category recommendation models. It was found that use of prefixsacn technique for pattern extraction has enhanced the leaning of random forest. Fig. 4 shows that Fuzzy normalization of input learning model has increases the prediction accuracy by 11.78% as compared to EDAMLP.

4. Conclusion

Business learns user behavior and it directly impacts on the turnover. So analysis of user done by various means. This work focus on shopping store purchase data and found that user do purchasing in a pattern and if product are arrange in same pattern then result will be more effective. In this work pattern extraction was done by PreFix Scan technique, extracted patterns were normalize by fuzzy min-max that enrich the learning data. Random forest model was trained by the normalize patterns to predict the purchase category prediction. Experiment was done on real dataset and result shows that proposed model has improved the prediction accuracy by 11.78%. In future scholar can use more deep data like social connection for increasing the prediction accuracy.

5. REFERENCES

1. S. Ojo et al., "Graph Neural Network for Smartphone Recommendation System: A Sentiment Analysis Approach for Smartphone Rating," in IEEE Access, vol. 11, pp. 140451-140463, 2023.
2. 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.
3. A. 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.
4. 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.
5. Chong Tat Chua, Hady W. Lauw, and Ee-Peng Lim. “Generative Models for Item Adoptions Using Social Correlation”. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 9, SEPTEMBER 2013.
6. 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.

7. MicolPolicarpo L, da Silveira DE, da Rosa RR, Antunes Stoffel R, da Costa CA, Victória Barbosa JL, Scorsatto R, Arcot T (2021) Machine learning through the lens of e-commerce initiatives: an up-to-date systematic literature review. Comput Sci Rev 41:100414.
8. N. Su, "Promoting by Looking Into Friend Circles: An Inadequately-Labeled and Socially-Aware Financial Technique for Products Recommendation," in IEEE Access, vol. 12, pp. 159833-159846, 2024.
9. R. Kwieciński, G. Melniczak and T. Górecki, "Comparison of Real-Time and Batch Job Recommendations," in IEEE Access, vol. 11, pp. 20553-20559, 2023.
10. C. Wang and P. Yue, "Learning Resource Recommendation Model Based on Collaborative Knowledge Graph Attention Networks," in IEEE Access, vol. 12, pp. 153232-153242, 2024.
11. Daniyal Irfan, Xuan Tang, Vipul Narayan, Pawan Kumar Mall, Swapnita Srivastava, V. Saravanan. "Prediction of Quality Food Sale in Mart Using the AI-Based TOR Method. 30 May 2022.
12. 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.
13. 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.
14. F. Busato and N. Bombieri, "On the Load Balancing Techniques for GPU Applications Based on Prefix-Scan," 2015 IEEE 9th International Symposium on Embedded Multicore/Many-core Systems-on-Chip, Turin, Italy, 2015, pp. 88-95.
15. Yile Ao, Hongqi Li, Liping Zhu, Sikandar Ali, Zhongguo Yang. "The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling". Journal of Petroleum Science and Engineering, Volume 174, 2019, Pages 776-789.
16. Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. The Stata Journal, 20(1), 3-29
17. https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset