

Exploring Ensemble Techniques For Market Basket Analysis

Nagarjuna Battula
Department of CSE
VFSTR Deemed to be University
Guntur, India
battulanagarjuna117@gmail.com

Sravanthi Kunamneni
Department of CSE
VFSTR Deemed to be University
Guntur, India
sravanthikunamneni5@gmail.com

Vamsi Puvvadi
Department of CSE
VFSTR Deemed to be University
Guntur, India
vamsipuvvadi123@gmail.com

Jahnavi Kommu
Department of CSE
VFSTR Deemed to be University
Guntur, India
jahnavikommu04@gmail.com

Syed Shareefunnisa
Asst. Professor, Department of CSE
VFSTR Deemed to be University
Guntur, India
syedshareefa@gmail.com

Abstract—Frequent pattern mining is essential to market basket analysis, which aims to identify meaningful relationships between products that consumers buy together. MBA is a form of analytical technique that marketers commonly use to understand their clients' purchasing behavior. In this research, we investigate two popular data mining strategies for finding frequent item sets from large transactional datasets: the Apriori and FP-growth algorithms. Furthermore, we use real-world datasets to compare these algorithms' performance in terms of scalability and runtime efficiency. The dataset's properties have a big impact on how well algorithms work. All things considered, this study advances knowledge of frequent pattern mining methods and how they affect market basket analysis, providing insightful advice for practitioners and researchers working in the retail analytics and data mining fields.

Index Terms—Apriori algorithm, FP-growth algorithm, Frequent Itemsets, Association rules

I. INTRODUCTION

Market Basket Analysis (MBA) is a key data mining technique that looks for significant correlations between items that customers typically purchase together. Targeted marketing methods are made easier and consumer behavior is better understood by marketers using this analytical approach. The foundation of MBA is the extraction of frequent itemsets, which represent items that frequently occur together in transactions. The FP-growth algorithm and the Apriori algorithm are two well-known algorithms that are frequently used in MBA for pattern mining.

An essential technique for mining frequent itemsets is the Apriori algorithm, which was put forth by Agrawal et al. in 1994. To create candidate itemsets and eliminate infrequent ones based on support—a metric for how frequently an itemset appears in the dataset—iterative processes are used. With large datasets, Apriori can become computationally expensive due to candidate generation and multiple database searches, despite its conceptual clarity and simplicity. The memory usage for

storing candidate itemsets is considerable. It ensures accuracy and comprehensiveness.

As an alternative, Han et al.'s 2000 introduction of the FP-growth algorithm provides a more effective method by creating a small data structure known as the FP-tree. By doing away with the necessity for candidate generation, this tree-based approach improves scalability and lowers computing overhead. Utilizing the FP-tree, FP-growth mines frequently occurring itemsets and generates association rules according to user-specified confidence and support levels. Because of the compact FP-tree structure, it uses less memory than Apriori. Both sparse and dense datasets can be used.

Market Basket Analysis with a Hash-Based Algorithm enhances understanding of consumer shopping patterns to optimize sales strategies in supermarkets[14]. In MBA, association rules are derived from frequent item sets, which are collections of items that regularly co-occur. These rules measure the strength and consistency of relationships between items and are typified by confidence and support measures. Whereas confidence assesses the conditional probability of one itemset occurring given the existence of another itemset, support measures the frequency of occurrence of an itemset.

II. RELATED WORK

Ahmad Ari Aldino et al.[1] provided the application of market basket analysis—which examines consumer purchasing patterns from sales transaction data—using FP-Growth and Apriori algorithms in the PDF. It contrasts the two algorithms' efficiency, demonstrating FP-Growth's superiority. The Universitas Teknokrat Indonesia study's authors highlight the value of data mining in assisting companies in making strategic decisions based on transaction data. Kitty S.Y. Chiu et al.[2] explains how Market basket analysis combines PCA and association rules to study consumer spending behavior to improve cross-selling strategies. By setting a threshold for PC correlations, this study revealed its impact on item set memory

and emphasized the importance of maintaining quality association rules. This study highlights the importance of PCA in identifying underlying patterns of consumer transactions by illuminating the interplay between underlying components and association rule analysis. This approach provides valuable insights into consumer preferences and purchasing decisions, ultimately optimizing market strategies to increase sales and customer satisfaction. Shish Kumar Dubey et al.[3] compares Market Basket Analysis with data mining methods to better understand customer preferences and improve sales tactics. It contrasts the application of Collaborative Filtering for product recommendations with the Apriori Algorithm for determining recurring purchasing trends. The study emphasizes how important it is to assess customer behavior and use data mining to analyze markets. The overall goal of the research is to enhance the processes involved in decision-making and maximize product recommendations derived from previous customer purchases. Behera Gayathri [4]introducing the FP-Bonsai algorithm for effective Market Basket Analysis in retailing is covered in the PDF. It demonstrates how FP-Bonsai gets around memory and I/O limitations to make frequent itemset discovery possible. The programme effectively performs market basket research on supermarket data, helping merchants better understand consumer purchasing patterns and develop their business plans. Maliha Hossain et al.[5]delves into the Apriori and FP Growth algorithms' performance evaluation for market basket analysis in the PDF. By concentrating on the best-selling products, it suggests cutting down on computation and contrasts the outcomes with and without product reduction. The results of the studies indicate that FP Growth performs better in terms of execution time than Apriori. The performance of the suggested approach in association rule mining is further demonstrated by the study's analysis of French Retail and Bakery Shop datasets. Mrignainy Kansal et al.[6]research provides the grocery product marketing strategy's use of Market Basket Analysis with the Eclat and Apriori algorithms . It contrasts the two methods' levels of efficiency, emphasizing Eclat's quicker speed and lower time-space complexity for small- to medium-sized datasets. The significance of lift values, support, and confidence in identifying relevant association rules is also emphasized in the paper. Sanket Sandip Khedkar et al.[7]explore FP-Tree and Apriori algorithms for market basket analysis in the PDF. It describes how these algorithms work, emphasising how they produce frequent item sets and determine support values. The significance of establishing a minimal threshold for association rule mining based on data observation is also mentioned in the document. S.Dinesh Kumar et al.[8] describes the use of market basket analysis to examine consumer buying trends in department shops. The document emphasises how crucial it is to determine the most popular item combinations in order to increase sales through tailored combo offers and discounts. The study highlights how popular pairings can be suggested to clients using Point of Sale (POS) data. Numerous tactics and methods are recommended to enhance the research of consumer behaviour and maximise marketing strategies, including

hereditary computations and KNIME Picture Handling Augmentation. YongmeiLiu and YongGuan [9]demonstrates focus on the effective mining of frequent item sets without candidate generation, and the PDF explains the FP-Growth algorithm for market basket analysis. It emphasizes how often item sets may be created using FP-Tree and stresses the significance of establishing minimal support and confidence standards. The study offers insights into the process of mining association rules and investigates the use of FP-Growth in managing huge databases. It also illustrates how the technique is really implemented in practice using program design tools like Visual C++. Abdul Rezha Efrat Najaf et al.[10]study delves into the FP-Growth method's use for Market Basket Analysis at XYZ Coffee Shop. The document describes the steps involved in the study approach, such as data collection, preparation, and evaluation. The study's main objective is to employ association rule mining to pinpoint customer purchase trends. The results show a relationship between product purchases and offer guidance for the coffee shop's marketing plans and product promotions. Warnia Nengsih [11]explains an comparison of Market Basket Analysis using and without using the Apriori algorithm in the PDF. It examines how rules are made, how support values are chosen, and how the outcomes are evaluated. The study highlights the significance of support and confidence values in rule generation by demonstrating that both approaches produce the same rules. Muhammad Raihan Pradana et al.[12]uses market basket analysis on retail sales data, and also the PDF explains how to apply the FP-Growth algorithm to forecast consumer behavior in a retailer that is experiencing a decline in sales of fashion products. A variety of marketing techniques, such as product bundling and discounts, have been used; the FP-Growth algorithm was selected because of its effectiveness. After testing 571 transaction data items, the study identified four rules with lift ratio values greater than 1, offering insights into product associations to improve sales tactics. Anshika Sharma and Himanshi Babbar [13]described that market Basket Analysis makes use of data mining techniques like Apriori and FP-Growth algorithms, which provide insights into customer behavior and suggest products. In order to properly handle and analyze data, the architecture of data mining incorporates a number of components. To analyze transactions, create marketing campaigns that work, and accurately identify security breaches, firms need to use these tactics. According to execution time, accuracy, and support, the study assesses how well the FP-Growth and Apriori algorithms perform. In general, data mining is essential for trend prediction, well-informed decision-making, and accurate and dependable results that improve company outcomes. Arini Wulandari et al.[14] investigates the application of Market Basket Analysis with a Hash-Based Algorithm to comprehend customer purchase behaviors in the PDF from the APICS conference in 2022. Popular goods such as butter, yogurt, and chocolate are identified by the study using frequent item sets and association criteria. The study highlights how crucial it is to forecast consumer demand and optimize inventory levels in response to their buying patterns. Supermarkets use KNIME

software to process data and find patterns in customer behavior and association rules. Liu Yongmei and Guan Yong [15] focus on the application of FP growth algorithm, which uses the FP tree structure to efficiently analyze frequent patterns, eliminating the need to generate candidate sets. Shopping cart analytics is critical for retailers to understand customer purchasing behavior and optimize sales strategies based on product relevance. For example, identifying correlations such as diaper and beer purchases can lead to strategic changes to the layout. This process involves creating a table of potential data items, constructing a frequent pattern tree, and concatenating them to obtain a set of frequent elements. Setting a minimum support threshold can help identify patterns that are important to retailer decision-making. Ching-Huang Yun et al. [16] covered an effective clustering approach for market basket data that makes use of the small-large (SL) ratio measurement in the PDF. For better clustering outcomes, the algorithm seeks to reduce the SL ratio in each group. With this new method, the algorithm shows notable speedups over previous studies in terms of execution time and produces excellent clustering results.

III. PROPOSED METHODOLOGY

The sequence of activities that are performed in this process involves that the transactional data from retail sources is pre-processed to remove irrelevant information and handle missing values. Frequent item sets are generated using algorithms like Apriori or FP-Growth or RandomForestClassifier or by ensembling any of the algorithms that have been mentioned. Association rules are then mined to reveal patterns of co-occurring items. Different algorithms are compared based on computational efficiency and pattern quality.

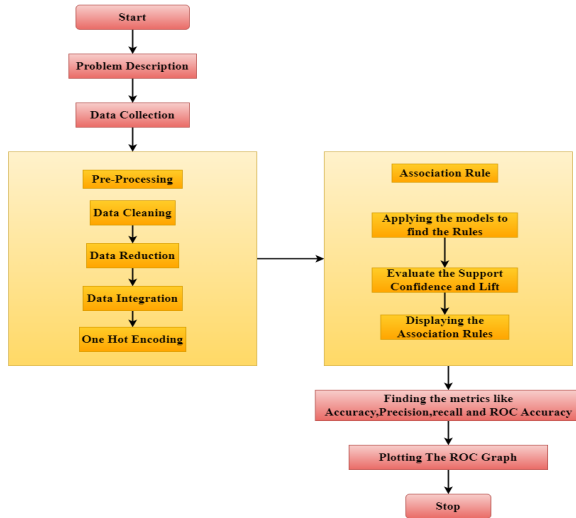


Fig. 1. Sequence of activities

A. Collection of data

The dataset utilized in this research paper was obtained from GitHub, focusing on retail transactions. The size of the

dataset that we have used is (522064, 7). It features several key attributes including BillNo, Itemname, Quantity, Date, Price, CustomerID, and Country. Each entry represents a transaction, detailing purchased items, their quantities, transaction dates, prices, and customer information. Spanning from December 2010 to December 2011, the dataset encompasses a diverse range of retail transactions.

B. Data Pre-processing

After the dataset was loaded, columns containing null values were displayed in a Pandas DataFrame. The mode was used in the 'Itemname' field to fill in any missing values. Then, 'BillNo' was used to combine transactions, putting together purchases made together. 'Itemname' underwent one-hot encoding, which transformed category data into binary representation. For compatibility, 'Itemname's' floating-point values were changed to strings. The outcome of this procedure was a binary representation of the transactions. By taking these precautions, data quality and compatibility with analysis techniques like pattern mining and market basket analysis are guaranteed.

- **Detecting and Handling the missing values in the dataset:** Using Python and Pandas for data preprocessing, the `isnull()` method helps find missing values. Then, to properly manage these null values, techniques like imputation, deletion, or interpolation are used. By guaranteeing data quality before analysis, this reduces bias and improves accuracy. Managing missing values correctly is essential to preserving the dataset's dependability and integrity during the analysis phase.

TABLE I
COLUMNS WITH MISSING VALUES

Column	Missing Values
Itemname	1455
CustomerID	134041

TABLE II
AFTER HANDLING THE MISSING VALUES

Column	Missing Values
BillNo	0
Itemname	0
Quantity	0
Date	0
Price	0
CustomerID	134041
Country	0

- **Data Consolidation:** It involves merging and organizing data from various sources into a cohesive format. This entails grouping related data based on common attributes and aggregating relevant information. Its purpose is to streamline data management, enhance quality, and enable

efficient analysis by presenting a unified view of the information. It eliminates redundancy, ensures consistency, and enhances accessibility for reporting and decision-making.

- **One-Hot Encoding:** Categorical variables are transformed into binary vectors for machine learning by one-hot encoding. All the elements in a category become 'cold' (0) and only one is 'hot' (1), forming binary vectors. Without ordinal interpretation, it guarantees that categorical variables are machine-readable. That might, however, result in a high-dimensional feature space. In general, it's an essential step in the preparation process for using categorical data in machine learning models.

C. Pattern Finding Algorithms

- 1) **Apriori:** Using a breadth-first search strategy and the Apriori principle, the Apriori algorithm is essential for quickly mining common itemsets. Because it makes correlations between products easier to extract, it is commonly utilized for activities like market basket analysis and product suggestions. Even with possible issues with sparse datasets or high dimensionality, its interpretability and scalability make it appropriate for a wide range of sectors. Apriori is still a useful tool for finding important patterns in transactional databases when it is implemented and optimized effectively.

Algorithm 1 Apriori Algorithm

```

1: Input: Dataset  $D$ , minimum support threshold  $minSup$ 
2: Output: Frequent itemsets
3: Initialize  $L_1$  with frequent 1-itemsets
4:  $k = 2$ 
5: while  $L_{k-1} \neq \emptyset$  do
6:   Generate candidate itemsets  $C_k$  from  $L_{k-1}$ 
7:   for all  $transaction \in D$  do
8:     for all  $candidate \in C_k$  do
9:       if  $candidate \subset transaction$  then
10:        Increment the count of  $candidate$ 
11:       end if
12:     end for
13:   end for
14:   Prune candidate itemsets  $C_k$  that do not meet  $minSup$ 
15:    $L_k = \{itemset \in C_k : support(itemset) \geq minSup\}$ 
16:    $k = k + 1$ 
17: end while
18: return Frequent itemsets  $\bigcup_k L_k$ 

```

- 2) **RandomForestClassifier:** RandomForestClassifier is an ensemble learning method for classification and regression. It constructs multiple decision trees, each trained on random subsets of the data and features. The final prediction is the mode of classification or mean prediction of the trees. Known for its robustness against overfitting and ability to handle high-dimensional data, it's widely applied in finance, healthcare, and marketing for tasks like classification and regression.

Algorithm 2 Random Forest Classifier

```

1: Input: Dataset  $D$ , number of trees  $N$ 
2: Output: Ensemble of decision trees  $RF$ 
3: for  $i = 1$  to  $N$  do
4:   Randomly select a subset  $D_i$  from  $D$  with replacement
5:   Train decision tree  $T_i$  on  $D_i$  using random feature subset
6:   Add  $T_i$  to the forest  $RF$ 
7: end for
8: return  $RF$ 

```

- 3) **FP-Growth:** The FP-growth approach creates compact FP-tree representations and iteratively mines patterns to efficiently find frequent itemsets in large datasets. Because of its two-pass methodology, it excels at tasks like online usage mining and market basket analysis. Its efficacy and scalability allow it to be applied in a variety of sectors and enable insights from large, complicated datasets. It maximizes cross-selling and product placement tactics in market basket analysis. It improves user experience and customizes content recommendations in web usage mining. It is useful for jobs like fraud detection, network optimization, and disease diagnosis in healthcare, banking, and telecommunications. All things considered, FP-growth enables analysts to extract practical insights, spurring innovation across a range of industries.

Algorithm 3 FP-growth Algorithm

```

1: Input: Dataset  $D$ , minimum support threshold  $minSup$ 
2: Output: Frequent itemsets
3: Construct the FP-tree  $T$  from  $D$ 
4: Initialize  $F$  as an empty set
5: for all  $item$  in  $T$  in reverse order of support do
6:   Initialize  $freqSet$  as  $\{item\}$ 
7:   Initialize  $conditionalPatternBase$  as empty
8:   for all  $path$  from  $item$  to the root in  $T$  do
9:     Add the path's support to  $conditionalPatternBase$ 
10:   end for
11:   Generate conditional FP-tree  $T'$  from  $conditionalPatternBase$ 
12:   if  $T'$  is not empty then
13:     Recursively mine  $T'$  for frequent itemsets with  $minSup$ 
14:     Add the frequent itemsets found in  $T'$  to  $F$ 
15:   end if
16: end for
17: return  $F$ 

```

- 4) **Adaboost:** AdaBoost, or Adaptive Boosting, is an ensemble learning technique that combines weak learners to form a strong classifier. It iteratively trains classifiers, focusing more on misclassified instances in each iteration. By assigning weights to classifiers based on their performance, AdaBoost creates a powerful clas-

sifier capable of handling complex classification tasks. Widely applied in face detection, object recognition, and bioinformatics, AdaBoost stands out for its versatility and effectiveness in boosting model performance.

Algorithm 4 AdaBoost Algorithm

- 1: **Input:** Dataset D , number of iterations T
 - 2: **Output:** Strong classifier $H(x)$
 - 3: Initialize instance weights $w_i = 1/N$, where N is the number of instances
 - 4: Initialize empty set $H(x)$ to store weak classifiers
 - 5: **for** $t = 1$ to T **do**
 - 6: Train weak learner $h_t(x)$ using weighted instances D and weights w
 - 7: Compute error ϵ_t of $h_t(x)$ on D
 - 8: Set classifier weight $\alpha_t = \frac{1}{2} \ln \left(\frac{1-\epsilon_t}{\epsilon_t} \right)$
 - 9: Update instance weights $w_i \leftarrow w_i \times \exp(-\alpha_t y_i h_t(x_i))$, where y_i is the true label
 - 10: Normalize weights $w_i \leftarrow \frac{w_i}{\sum_{i=1}^N w_i}$
 - 11: Add weak classifier $h_t(x)$ with weight α_t to $H(x)$
 - 12: **end for**
 - 13: **return** Strong classifier $H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$
-

- 5) **xgboost:** XGBoost, or Extreme Gradient Boosting, is a high-performance implementation of gradient boosting machines. Renowned for its efficiency and scalability, it sequentially builds a tree ensemble model to correct errors from previous trees. With built-in regularization and handling of missing data, XGBoost excels in classification, regression, and ranking tasks. Its speed, flexibility, and ability to achieve state-of-the-art results across various domains have made it a preferred choice in machine learning competitions and real-world applications.

Algorithm 5 XGBoost Algorithm

- 1: **Input:** Training data D , number of iterations N , hyperparameters
 - 2: **Output:** Ensemble of boosted trees G
 - 3: Initialize ensemble G as an empty set
 - 4: **for** $i = 1$ to N **do**
 - 5: Fit tree T_i to the residuals of D using hyperparameters
 - 6: Update G by adding T_i with a weight determined by performance
 - 7: **end for**
 - 8: **return** G
-

IV. EXPERIMENTAL RESULTS

Presented below are the experimental results showcasing the effectiveness of various ensemble techniques in Market Basket Analysis. Each algorithm's performance metrics are summarized in the table.

TABLE III
PERFORMANCE METRICS

Algorithm	Accuracy	Precision	Recall	F1 Score	ROC AUC
Apriori	0.5025	0.5027	0.5045	0.5036	0.5025
FP-growth	0.5025	0.5027	0.5045	0.5036	0.5025
AdaBoost+FPgrowth	0.5268	0.5898	0.1774	0.2727	0.5269
XGBoost+FPgrowth	0.6739	0.6377	0.8059	0.7120	0.6738
RandomForest+FPgrowth	0.9356	0.9029	0.9761	0.9381	0.9355

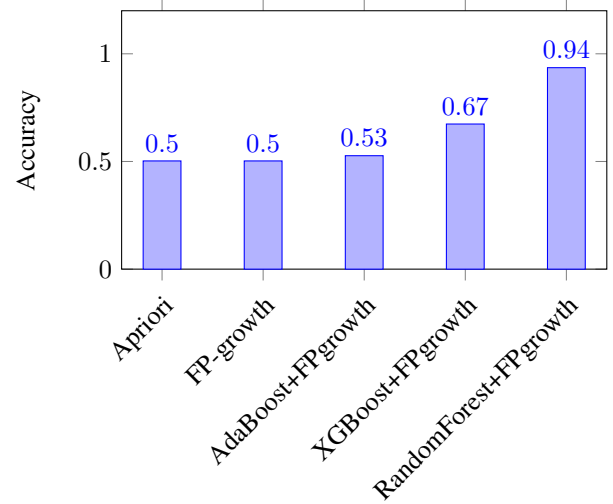


Fig. 2. Accuracy for models

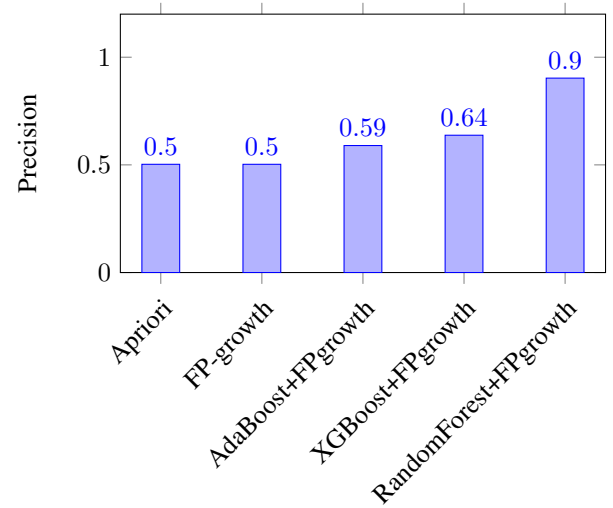


Fig. 3. Precision for models

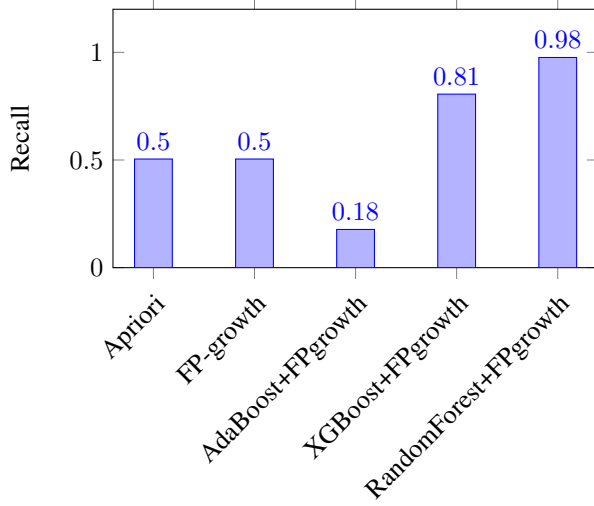


Fig. 4. Recall for models

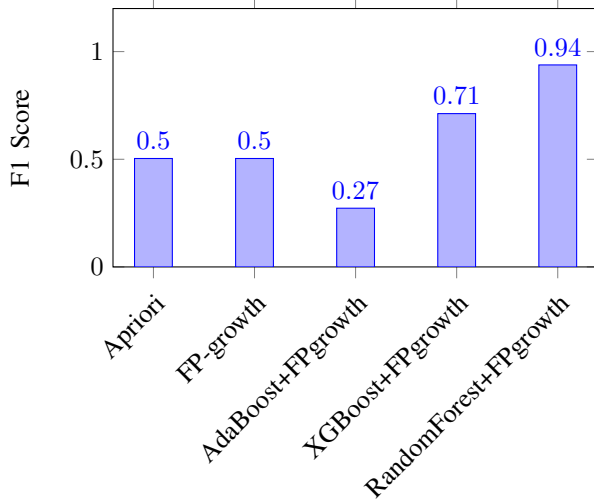


Fig. 5. F1 Score for models

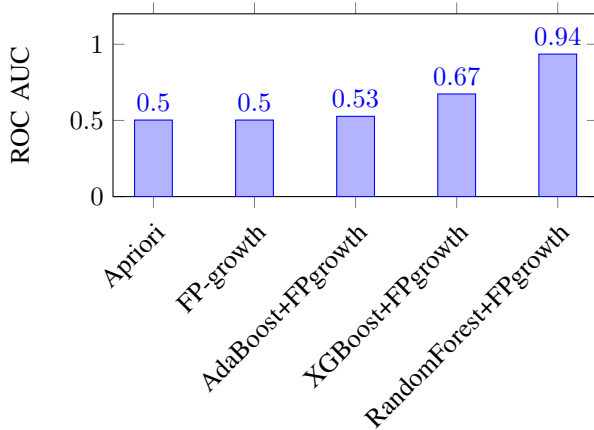


Fig. 6. ROC AUC for models

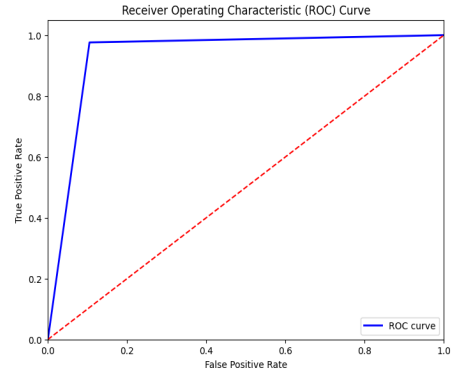


Fig. 7. ROC Curve for RandomForestClassifier with FP-Growth

V. CONCLUSION

The Apriori and FP-growth algorithms exhibit comparable performance across all metrics, with accuracy, precision, recall, F1 score, and ROC AUC around 0.50. This indicates similar effectiveness in frequent pattern mining for market basket analysis. However, the AdaBoost with FP-growth model shows a lower recall (0.1774) and F1 score (0.2727), yet higher precision (0.5898), implying a trade-off between false positives and true positives. Conversely, the XGBoost with FP-growth model demonstrates improved performance with higher accuracy (0.6739), recall (0.8059), F1 score (0.7120), and ROC AUC (0.6738), suggesting a better balance between precision and recall. Remarkably, the RandomForest with FP-growth model excels in all metrics, achieving the highest accuracy (0.9356), precision (0.9029), recall (0.9761), F1 score (0.9381), and ROC AUC (0.9355). This highlights the efficacy of combining Random Forest with FP-growth for superior market basket analysis.

From the ROC curve analysis of the ensemble model, combining RandomForestClassifier with FP-growth, several conclusions emerge:

- The blue curve depicts the ensemble model's ROC curve, surpassing the diagonal line, indicating superiority over random classification.
- Initially steep, the ROC curve reflects high sensitivity with low false positive rates, showcasing the ensemble model's effective identification of positive instances.
- However, with increasing false positive rates, the true positive rate rises more slowly, indicating a gradual increase in false positive predictions.
- The substantial area under the curve underscores the ensemble model's strong performance in distinguishing between positive and negative instances.

Finally, when compared to other models, the ensemble model that combines FP-growth and RandomForestClassifier performs better. It performs better than other methods and improves predictive capability by effectively utilizing both algorithms.

REFERENCES

- [1] Ahmad Ari Aldino et al. "Comparison of market basket analysis to determine consumer purchasing patterns using fp-growth and apriori algorithm". In: *2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*. IEEE. 2021, pp. 29–34.
- [2] Kitty SY Chiu et al. "Market-basket analysis with principal component analysis: an exploration". In: *IEEE International Conference on Systems, Man and Cybernetics*. Vol. 3. IEEE. 2002, 6–pp.
- [3] Shish Kumar Dubey et al. "Comparative analysis of market basket analysis through data mining techniques". In: *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. IEEE. 2021, pp. 239–243.
- [4] Behera Gayathri. "Efficient market basket analysis based on FP-Bonsai". In: *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE. 2017, pp. 788–792.
- [5] Maliha Hossain, AHM Sarowar Sattar, and Mahit Kumar Paul. "Market basket analysis using apriori and FP growth algorithm". In: *2019 22nd international conference on computer and information technology (ICCIT)*. IEEE. 2019, pp. 1–6.
- [6] Mrignainy Kansal et al. "Implementing Market Basket Analysis Using Eclat And Apriori Algorithm On Grocery Product Marketing Strategy". In: *2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE)*. IEEE. 2023, pp. 465–469.
- [7] Sanket Sandip Khedkar and Sangeeta Kumari. "Market basket analysis using A-priori algorithm and FP-tree algorithm". In: *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*. IEEE. 2021, pp. 1–6.
- [8] S Dinesh Kumar, K Soundarapandiyar, and S Meera. "Commiserating Customers' Purchasing Pattern Using Market Basket Analysis". In: *2022 1st International Conference on Computational Science and Technology (ICCSST)*. IEEE. 2022, pp. 1–4.
- [9] Yongmei Liu and Yong Guan. "Fp-growth algorithm for application in research of market basket analysis". In: *2008 IEEE International Conference on Computational Cybernetics*. IEEE. 2008, pp. 269–272.
- [10] Abdul Rezha Efrat Najaf et al. "Implementation Frequent Pattern Growth (FP-Growth) Algorithm on Market Basket Analysis for Determining Consumer Purchasing Patterns (Case Study: XYZ Coffee Shop)". In: *2023 IEEE 9th Information Technology International Seminar (ITIS)*. 2023, pp. 1–5. DOI: 10.1109/ITIS59651.2023.10420163.
- [11] Warnia Nengsih. "A comparative study on market basket analysis and apriori association technique". In: *2015 3rd international conference on information and communication technology (ICoICT)*. IEEE. 2015, pp. 461–464.
- [12] Muhammad Raihan Pradana et al. "Market Basket Analysis Using FP-Growth Algorithm On Retail Sales Data". In: *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. IEEE. 2022, pp. 86–89.
- [13] Anshika Sharma and Himanshi Babbar. "Analysis of Data Mining Algorithms in Market Basket Analysis". In: *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)*. IEEE. 2023, pp. 275–280.
- [14] Arini Wulandari et al. "Analysis of Customer Expense using Market Basket Analysis Model with Hash-Based Algorithm and Association Rules". In: *2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS)*. IEEE. 2022, pp. 10–14.
- [15] Liu Yongmei and Guan Yong. "Application in market basket research based on FP-growth algorithm". In: *2009 WRI World Congress on Computer Science and Information Engineering*. Vol. 4. IEEE. 2009, pp. 112–115.
- [16] Ching-Huang Yun, Kun-Ta Chuang, and Ming-Syan Chen. "An efficient clustering algorithm for market basket data based on small large ratios". In: *25th Annual International Computer Software and Applications Conference. COMPSAC 2001*. IEEE. 2001, pp. 505–510.