Analysis of Apriori and FP-Growth Algorithms for Market Basket Insights: A Case Study of The Bread Basket Bakery Sales

Hery[1,*], Andree E. Widjaja[2]

[1,2]Department of Information System, Universitas Pelita Harapan, Tangerang 15811, Indonesia

ABSTRACT

Market basket analysis is a crucial technique in retail for uncovering associations between items frequently purchased together. This study aims to compare the effectiveness of the Apriori and FP-Growth algorithms using sales data from "The Bread Basket" bakery, comprising 20,507 transactions. Key variables include TransactionNo, Items, DateTime, Daypart, and DayType. The data underwent preprocessing steps, including cleaning, tokenization, and feature extraction using TF-IDF. The Apriori and FP-Growth algorithms were implemented with hyperparameter tuning and an 80/20 training/testing split. Performance metrics were evaluated, revealing that Apriori had an execution time of 4.08 seconds and memory usage of 45.36 MiB, whereas FP-Growth exhibited an execution time of 4.15 seconds and significantly lower memory usage at 0.08 MiB. The quality of the association rules was assessed by metrics such as support, confidence, and lift. For example, the Apriori algorithm generated the rule {Alfajores} → {Coffee} with support 0.018885, confidence 0.520000, and lift 1.087090, while FP-Growth produced the rule {Scone} → {Coffee} with support 0.017829, confidence 0.519231, and lift 1.085482. FP-Growth generally outperformed Apriori, particularly in memory efficiency, due to its use of the FP-tree data structure, which reduces the need for multiple database scans. The practical implications for "The Bread Basket" bakery include optimizing product placement and inventory management based on the identified associations, such as placing Coffee near Cake or Medialuna to encourage complementary purchases. The study concludes that while both algorithms effectively generate meaningful association rules, FP-Growth's superior memory efficiency makes it more suitable for large datasets. Limitations include data quality and the study's scope, confined to a single bakery. Future research should explore hybrid approaches, real-time data analysis, and applications across different retail sectors to enhance market basket analysis techniques further.

Keywords Retail Data Mining, Apriori, FP-Growth, Market Basket Analysis, Association Rules

Submitted 10 January 2024
Accepted 20 April 2024
Published 1 June 2024

Corresponding author
Hery, hery.fik@uph.edu

Additional Information and Declarations can be found on page 81

DOI: 10.47738/jdmdc.v1i1.2

INTRODUCTION

Data mining, often called knowledge discovery in databases (KDD), is extracting valuable information and patterns from large datasets. This interdisciplinary subfield of computer science involves using various techniques from statistics, machine learning, and database systems to identify patterns, anomalies, and correlations within large datasets. The primary goal of data mining is to transform raw data into meaningful and actionable insights that can aid decision-making processes across different domains. Data mining plays a pivotal role in the digital landscape dominated by e-commerce. E-commerce platforms generate massive amounts of data daily, including customer purchase histories, browsing patterns, product reviews, and transaction records. Data

How to cite this article: Hery and A. E. Widjaja, "Analysis of Apriori and FP-Growth Algorithms for Market Basket Insights: A Case Study of The Bread Basket Bakery Sales," *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 1, pp. 63-83, 2024.

mining techniques enable businesses to analyze this data to uncover hidden patterns and trends that can inform a wide range of business decisions.

The application of data mining techniques extends to fields such as higher education, which involves extracting relevant information from databases to facilitate decision-making processes. In business intelligence, data mining is utilized to discover trends, patterns, and relationships within vast datasets, aiding in making informed business decisions [1]. Data mining methodologies are increasingly integrated into customer relationship management (CRM) applications, with classification models being vital tools in practice [2]. In conclusion, data mining techniques are powerful tools for businesses to extract valuable insights from vast datasets, enabling informed decision-making across various sectors. By uncovering hidden patterns, relationships, and trends within data, businesses can enhance their strategic decision-making processes and gain a competitive edge in today's data-driven landscape.

For instance, by analyzing customer purchase histories, e-commerce businesses can develop personalized product recommendations, enhance customer engagement, and ultimately increase sales. Moreover, data mining aids in identifying customer segments, understanding consumer behavior, and predicting future purchasing trends. By leveraging these insights, e-commerce companies can tailor their marketing strategies to target specific customer groups more effectively, optimize pricing strategies, and manage inventory levels efficiently. The ability to analyze and interpret vast amounts of data in real-time has become a critical competitive advantage in the highly dynamic e-commerce landscape.

Data mining facilitates the identification of sales patterns and trends, enabling retailers to make informed decisions about product placement, store layout, and inventory management. For example, association rule mining can reveal which products are frequently purchased together, allowing retailers to optimize product bundling and cross-selling strategies. Predictive analytics can forecast demand for specific products, helping retailers maintain optimal stock levels and reduce the risk of overstocking or stockouts.

Furthermore, data mining supports CRM by providing insights into customer loyalty and satisfaction. Retailers can use these insights to design targeted marketing campaigns, personalize customer interactions, and enhance the overall shopping experience. The strategic application of data mining in retail has proven to be a game-changer, driving increased sales, improved customer retention, and enhanced operational efficiency.

Data mining has significantly impacted the retail sector by enabling businesses to extract valuable insights from vast datasets, leading to improved decision-making processes and enhanced customer experiences. Customer classification in the Indian retail sector has been enhanced by applying various machine learning approaches, allowing for a deeper understanding of consumer behavior and preferences [3]. Market basket analysis, a key data mining technique, has been instrumental in e-retailing, facilitating cross-selling, market basket analysis, and customer retention strategies [4]. Moreover, big data analysis in retail business processes has revolutionized store layout and sales bundling, providing retailers with valuable insights for optimizing their operations [5].

The concept of experiential retailing leveraged by data analytics has gained prominence, emphasizing the importance of data-driven decision-making in enhancing in-store customer experiences and driving retail success [6]. Retail analytics, particularly store segmentation based on rule-based purchasing behavior analysis, has emerged as a critical tool for retailers to understand customer preferences and tailor their strategies accordingly [7]. Product sales promotion utilizing purchase pattern analysis through the FP-Growth Algorithm, have enabled retailers to optimize their promotional efforts and inventory management, leading to increased sales and customer engagement [8]. Data mining applications in the retail sector have transformed traditional business practices by providing retailers with actionable insights derived from complex datasets. These insights have empowered retailers to enhance customer segmentation, personalize marketing strategies, optimize inventory management, and detect fraudulent activities, ultimately leading to improved operational efficiency and sustainable business growth.

Market basket analysis, a fundamental data mining technique, plays a crucial role in identifying associations and relationships between items frequently purchased together in retail settings. This method examines basket data to uncover valuable insights into customers' purchase intentions and behaviors [4]. By determining the associations between entities, market basket analysis enables a deeper understanding of the relationships between products within shopping baskets, aiding in strategic decision-making processes [9]. In a case study focusing on e-retailing, market basket analysis was conducted on a dataset comprising transactions involving two or more product categories, highlighting the significance of this technique in understanding consumer behavior and preferences [10]. Market basket analysis aims to identify products commonly purchased together by customers, providing retailers with valuable information for optimizing product placement and marketing strategies [11]. Market basket analysis, or association rule learning or affinity analysis, is a versatile data mining technique applicable across various fields, including marketing, bioinformatics, and education [12].

This technique is particularly useful in the retail industry for understanding customer purchasing behavior and uncovering hidden patterns within transaction data. The primary objective of market basket analysis is to discover itemsets that frequently co-occur in customer transactions, thereby providing insights into product affinities and customer preferences.

Several techniques are employed in market basket analysis, with association rule mining being the most prominent. Association rule mining involves identifying frequent itemsets and generating rules that describe the likelihood of items being purchased together. Key techniques in association rule mining include the Apriori algorithm and the FP-Growth algorithm. The Apriori algorithm works by iteratively identifying frequent itemsets and using these itemsets to generate association rules. It relies on a breadth-first search strategy and the property of downward closure, which states that if an itemset is frequent, all its subsets must also be frequent. This algorithm is effective but can be computationally expensive for large datasets. The FP-Growth algorithm, on the other hand, uses a depth-first search strategy and a compressed data structure called the frequent pattern tree (FP-tree) to mine frequent itemsets without candidate generation. This approach is more memory-efficient and faster than the Apriori algorithm, especially for large datasets with numerous transactions.

Market basket analysis is crucial for understanding customer behavior because it reveals the underlying relationships between products customers purchase. By analyzing these relationships, retailers can gain insights into customer preferences, shopping habits, and product affinities. This information is invaluable for designing effective marketing strategies, optimizing product placement, and enhancing the shopping experience.

Market basket analysis provides numerous benefits for retailers in understanding consumer behavior, optimizing product placement, and enhancing marketing strategies. Market basket analysis allows retailers to identify associations and relationships between frequently purchased products by analyzing transactional data, providing valuable insights into customer preferences and purchase patterns [13]. This analysis enables retailers to understand customer behavior more deeply, effectively tailor their product offerings and promotions to meet consumer needs[9]. Market basket analysis helps retailers uncover hidden patterns and trends in customer purchasing behavior by extracting associations or co-occurrences from transactional databases, facilitating targeted marketing strategies and personalized customer experiences [14].

The sales data from "The Bread Basket" store provides a comprehensive record of customer transactions over an extended period. This dataset includes detailed information about each transaction, capturing various aspects of the purchasing behavior of the bakery's customers. The data encompasses a wide range of products and reflects the diverse purchasing patterns of the bakery's clientele. This study's primary objective is to comprehensively compare the Apriori and FP-Growth algorithms in the context of market basket analysis. By applying both algorithms to the bakery sales data from "The Bread Basket," the study aims to evaluate their performance, effectiveness, and suitability for uncovering meaningful associations in transactional data.

Determining the effectiveness of the Apriori and FP-Growth algorithms is crucial for several reasons. First, performance evaluation is essential to understand the strengths and limitations of each algorithm, which helps in selecting the most appropriate method for different types of datasets and business contexts. This evaluation includes factors such as execution time, memory usage, and the quality of the generated rules. Second, the study's practical implications provide valuable insights for businesses looking to implement market basket analysis. Businesses can optimize their marketing strategies, improve product offerings, and enhance customer satisfaction by identifying the algorithm that yields the most actionable insights. Lastly, this comparative study contributes to the broader field of data mining by providing empirical evidence on the performance of two widely used algorithms. It adds to the existing body of knowledge and offers a reference point for future research and algorithm development.

The expected outcomes and contributions of the study are multifaceted. First, a comprehensive comparison is anticipated, providing a detailed analysis of the Apriori and FP-Growth algorithms, highlighting their relative strengths and weaknesses. This comparison will be based on various performance metrics, including execution time, memory efficiency, and the quality of the association rules generated. Second, the study analyzes the bakery sales data to uncover valuable insights into customer purchasing behavior. These insights can inform strategic decisions for "The Bread Basket" and similar businesses, improving performance and customer satisfaction. Lastly, the results of this study will offer practical guidance for data scientists, business analysts, and retail managers on selecting the most suitable algorithm for market basket analysis. The study will recommend when and how to use each algorithm based on the dataset's characteristics and the business objectives. In conclusion, this research aims to enhance the understanding of the Apriori and FP-Growth algorithms' capabilities in market basket analysis, providing both theoretical contributions and practical applications for improving business strategies in the retail industry.

Literature Review

Market Basket Analysis

Market basket analysis (MBA) is a data mining technique to identify associations between items purchased together in transactions. This method aims to uncover patterns in large datasets to understand customer purchasing behavior better. Historically, MBA has evolved alongside advances in data mining and business intelligence. Initially, it gained traction in the 1990s with the development of association rule mining techniques. Retailers quickly adopted MBA to leverage transactional data, transforming their approach to inventory management, marketing strategies, and customer relationship management.

The benefits of MBA for retailers are manifold. Firstly, it enhances inventory management by revealing which products are frequently bought together, allowing retailers to optimize stock levels, reduce stockouts, and minimize excess inventory [15]. Secondly, it improves marketing strategies by enabling the creation of targeted promotions and product bundles, which boost the effectiveness of cross-selling and upselling efforts. Thirdly, MBA helps design store layouts that encourage additional purchases, increasing overall sales. Lastly, by offering personalized recommendations based on purchase patterns, MBA enhances the customer experience, fostering satisfaction and loyalty. Successful implementations of MBA in retail include Walmart's use of transaction data analysis for inventory and supply chain optimization, Amazon's recommendation engine that drives significant sales through personalized suggestions, and Tesco's loyalty program data analysis to tailor promotions and product assortments.

Several key concepts and terminologies underpin market basket analysis. An itemset is a collection of one or more items bought together in a transaction. Support is the proportion of transactions in the dataset that contain a particular itemset, indicating its frequency. Confidence measures the likelihood that a transaction containing a specific item also includes another item, expressed as a conditional probability. Lift assesses the strength of an association rule by comparing the observed support to the expected support if the items were independent [16].

Frequent itemsets are groups of items that appear together in a dataset with a frequency above a predefined threshold. Identifying these itemsets is crucial for generating meaningful association rules, which reveal significant patterns in customer purchasing behavior. Association rules are "if-then" statements that describe the relationship between items in frequent itemsets, consisting of an antecedent (if part) and a consequent (then part). For example, an association rule might state, "If a customer buys bread, they are likely to buy butter."

The evaluation of association rules uses several metrics. Support measures the frequency of the itemset in the dataset, with higher support indicating a more common occurrence. Confidence assesses the reliability of the rule, with higher confidence indicating a stronger association between the items. Lift evaluates the strength of the rule by comparing the observed co-occurrence of items to their expected co-occurrence under independence. A lift value greater than 1 suggests a positive association, while a value less than 1 indicates a negative association [17].

In summary, market basket analysis is a powerful tool that provides retailers with valuable insights into customer behavior and market trends. Understanding key concepts and applying appropriate metrics enables retailers to harness the full potential of MBA, enhance their business strategies, and improve customer satisfaction.

Association Rule Mining

Association rule mining is a fundamental concept in data mining that aims to discover interesting relationships, correlations, or associations among a large set of data items. This technique helps identify patterns within transaction data, revealing significant insights about customer behavior, product affinities, and market trends [18].

The process of discovering relationships between items involves several key steps. Initially, data is collected and preprocessed to ensure it is suitable for analysis. The next step is the identification of frequent itemsets, which are groups of items that appear together frequently in transactions. From these frequent itemsets, association rules are generated in the form of "if-then" statements that describe how the presence of one item in a transaction implies the presence of another item. These rules are then evaluated using various metrics to determine their strength and reliability [19].

Association rule mining has a wide range of applications across various industries. Retail analyzes customer purchase data to optimize product placement, design effective promotions, and improve inventory management. In e-commerce, it powers recommendation systems that suggest products to customers based on their browsing and purchasing history. In healthcare, it helps discover relationships between medical conditions and treatments to improve patient care. In telecommunications, it identifies patterns in customer usage data to enhance service offerings. In finance, it detects fraudulent activities by analyzing transaction patterns.

Frequent itemsets are central to association rule mining. They are sets of items that appear together frequently in a dataset. Identifying frequent itemsets is crucial because they form the basis for generating association rules. With frequent itemsets, it would be easier to derive meaningful and actionable insights from the data. Several methods exist for identifying frequent itemsets. The most well-known method is the Apriori algorithm, which uses a bottom-up approach to generate candidate itemsets and then prunes those not meeting the minimum support threshold. Another method is the FP-Growth algorithm, which uses a compressed data structure called an FP-Tree to mine frequent patterns without candidate generation directly. Both methods aim to efficiently identify frequent itemsets in the dataset, minimizing computational overhead [20].

Frequent itemsets significantly impact the generation of association rules. Once frequent itemsets are identified, association rules can be derived by evaluating the confidence of potential rules formed from these itemsets. Rules that meet the minimum confidence threshold are considered strong associations and are included in the final set of rules. The quality of the generated rules heavily depends on the accuracy of the identified frequent itemsets.

Discovering frequent itemsets offers numerous benefits for market basket analysis. For retailers, it enables the development of effective cross-selling and up-selling strategies by identifying commonly purchased products. It also helps optimize inventory management by predicting demand for specific product combinations, reducing stockouts and overstock situations. Additionally, frequent itemsets provide insights into customer behavior, allowing retailers to design personalized promotions and improve the shopping experience. Businesses can make data-driven decisions that enhance their competitive edge and drive growth by leveraging frequent itemsets.

In conclusion, association rule mining and the discovery of frequent itemsets are powerful techniques that provide valuable insights into customer behavior and market trends. These insights can significantly enhance business strategies and operations across various industries, improving performance and competitive advantage.

Apriori Algorithm

The Apriori algorithm is a foundational method in data mining used to identify frequent itemsets and generate association rules. Developed by Rakesh Agrawal and Ramakrishnan Srikant in 1994, the algorithm operates on the principle that any subset of a frequent itemset must also be frequent. This property, known as the downward closure property or anti monotonicity, allows the algorithm to efficiently prune the search space by focusing only on candidate itemsets that have the potential to be frequent.

The process of the Apriori algorithm begins with the identification of all individual items in the dataset and their support counts. Items that meet the minimum support threshold are selected to form the set of frequent 1-itemsets (L1). In the next step, candidate k-itemsets (Ck) are generated by joining frequent (k-1)-itemsets (Lk-1) with themselves. The algorithm uses the apriori property to prune candidate itemsets that contain infrequent (k-1)-itemsets. The database is then scanned to count the support of each candidate k-itemset, and those meeting the minimum support threshold form the set of frequent k-itemsets (Lk). This candidate generation and support counting process is repeated until no new frequent itemsets are found. Finally, association rules are generated from the frequent itemsets by calculating confidence for each potential rule and selecting those that meet the minimum confidence threshold.

For example, consider a dataset with transactions such as {Bread, Milk}, {Bread, Diaper, Beer, Eggs}, {Milk, Diaper, Beer, Coke}, {Bread, Milk, Diaper, Beer}, and {Bread, Milk, Diaper, Coke}. The Apriori algorithm would first identify frequent 1-itemsets like Bread, Milk, Diaper, and Beer. It would then generate candidate 2-itemsets such as {Bread, Milk} and {Bread, Diaper}, count their support, and prune those that do not meet the minimum support threshold. This process continues iteratively, generating larger itemsets and deriving association rules from these frequent itemsets.

The strengths of the Apriori algorithm include its simplicity and effectiveness in discovering frequent itemsets and generating association rules. It provides a foundational approach for many other algorithms and techniques in association rule mining. However, the Apriori algorithm also has limitations, such as computational complexity and significant memory usage due to multiple database scans and the generation of many candidate itemsets. Its performance can degrade rapidly as the dataset size and the number of items increase.

The Apriori algorithm is most effective with small to medium-sized datasets with manageable computational complexity and memory usage. It performs well with low-dimensional data, where the number of unique items is relatively low, and is suitable for exploratory data analysis where generating initial insights and patterns is more important than computational efficiency.

FP-Growth does not generate candidate itemsets and is more efficient with large datasets than other association rule mining algorithms. FP-Growth uses a compact data structure (FP-Tree) to mine frequent patterns directly. The ECLAT algorithm, which uses a vertical data format, is more efficient in certain scenarios and avoids generating candidate itemsets explicitly, making it suitable for dense datasets. Despite its limitations, the Apriori algorithm remains a valuable tool for exploratory analysis and has paved the way for developing more advanced algorithms like FP-Growth and ECLAT.

FP-Growth Algorithm

The FP-Growth (Frequent Pattern Growth) algorithm is a highly efficient method for mining frequent itemsets without the need for candidate generation, addressing the limitations of the Apriori algorithm. FP-Growth utilizes a compact data structure called the FP-Tree to represent the dataset. This approach significantly reduces the computational cost and memory usage, making it particularly effective for large datasets.

The process of the FP-Growth algorithm begins with the construction of the FP-Tree. The first step involves scanning the database to identify all frequent items and their support counts. Items are then sorted in descending order of frequency, and infrequent items are removed. The FP-Tree is built by iterating through the transactions, adding each transaction's items as branches to the tree and incrementing the counts of existing nodes if they already exist. Once the FP-Tree is constructed, the algorithm mines the tree for frequent patterns. This involves generating conditional FP-Trees for each item, sub-databases consisting of the set of prefix paths that co-occur with the item. The FP-Growth algorithm is then applied recursively to each conditional FP-Tree to find frequent patterns. Finally, the frequent items generated from the conditional FP-Trees are combined with the prefix items to form the final set of frequent itemsets.

For example, consider a dataset with transactions such as {Bread, Milk}, {Bread, Diaper, Beer, Eggs}, {Milk, Diaper, Beer, Coke}, {Bread, Milk, Diaper, Beer}, and {Bread, Milk, Diaper, Coke}. The FP-Growth algorithm would first scan the database to determine the frequency of individual items, then construct the FP-

Tree by sorting items in each transaction by frequency and adding them to the tree. The algorithm would then mine the FP-Tree to generate frequent itemsets, such as {Bread, Milk} and {Diaper, Beer}.

The strengths of the FP-Growth algorithm include its efficiency and scalability. FP-Growth significantly reduces the number of database scans and the need for candidate generation, making it faster and more efficient than the Apriori algorithm, especially for large datasets. The compact FP-Tree data structure also saves memory and computational resources, allowing FP-Growth to handle dense datasets with numerous transactions. However, the algorithm does have some limitations. Constructing the FP-Tree can be complex and time-consuming, particularly for very large datasets with many unique items. Additionally, extracting conditional pattern bases and constructing conditional FP-Trees can be computationally intensive. FP-Growth remains a highly effective algorithm for mining frequent itemsets in large datasets despite these challenges.

The FP-Growth algorithm is most effective for large datasets, where the Apriori algorithm would be inefficient due to the large number of candidates itemsets. It performs well with dense datasets that have many frequent itemsets and transactions with numerous items. Due to its efficiency, FP-Growth is also suitable for applications requiring real-time analysis and pattern discovery.

Compared with other association rule mining algorithms, FP-Growth outperforms Apriori by eliminating the need for multiple database scans and candidate generation. The ECLAT (Equivalence Class Clustering and bottom-up Lattice Traversal) algorithm, which uses a vertical data format, is more efficient than Apriori in certain scenarios and avoids explicitly generating candidate itemsets. However, FP-Growth generally performs better with large and dense datasets.

In conclusion, the FP-Growth algorithm offers a powerful and efficient method for mining frequent itemsets and generating association rules. Its compact data structure and recursive pattern growth approach make it highly suitable for large and dense datasets. Despite some challenges in tree construction and memory consumption, FP-Growth remains a preferred choice for many real-world applications in association rule mining.

Comparative Studies

The literature comparing the Apriori and FP-Growth algorithms provides valuable insights into their relative performance and applicability. Key studies, such as the seminal work by [4], which introduced the FP-Growth algorithm, demonstrate that FP-Growth significantly outperforms Apriori in terms of speed and memory efficiency, especially for large datasets. Agrawal and Srikant's original paper on the Apriori algorithm in 1994 laid the foundation for association rule mining, and subsequent studies have built on this work to evaluate the performance of Apriori against newer algorithms like FP-Growth. [5] provided a comprehensive comparison of various frequent itemset mining algorithms, highlighting the strengths and weaknesses of each algorithm in different contexts. [6] focused on the implementation efficiency of Apriori and FP-Growth, concluding that FP-Growth is superior in terms of execution time and memory usage for dense and large datasets.

Analysis of findings from previous research consistently shows that FP-Growth is more efficient and scalable than Apriori, due to its compact data structure and elimination of candidate generation. Performance metrics commonly used in these comparative studies include execution time, memory usage, number of candidates itemsets generated, scalability, and the accuracy and quality of the generated rules. The insights gained from these comparisons indicate that while Apriori is simpler and more straightforward to implement, FP-Growth is better suited for large-scale data mining applications.

Despite the extensive research on Apriori and FP-Growth, there are gaps in the existing literature, particularly in their application to specific industries such as bakery sales. Most studies focus on general contexts, leaving a gap in industry-specific applications. Additionally, there is limited research on the real-time analysis capabilities of these algorithms, which is crucial for dynamic industries like retail. Furthermore, few studies delve deeply into how these algorithms can uncover nuanced customer behavior patterns specific to bakery sales.

Further research is needed to address these gaps and provide tailored insights for the bakery industry. Understanding the specific benefits and limitations of Apriori and FP-Growth in the context of bakery sales can help businesses optimize product offerings and marketing strategies. The current study aims to fill these gaps by providing a detailed comparison of Apriori and FP-Growth specifically for bakery sales. This research will offer practical recommendations for bakeries on leveraging these algorithms to enhance their business strategies and customer satisfaction.

Future research directions include investigating the integration of Apriori and FP-Growth with real-time data processing systems to enable dynamic and responsive market basket analysis. Exploring hybrid approaches that combine the strengths of both algorithms can maximize efficiency and accuracy. Another important direction is developing and applying advanced metrics to evaluate the quality and impact of association rules in the context of bakery sales. Additionally, extending research to include customer segmentation and personalized marketing strategies based on insights from market basket analysis can further enhance business performance. Conducting comparative studies across different retail sectors will also help generalize findings and develop industry-specific best practices.

Method

The methodology of this study is visually represented through a comprehensive flowchart, which includes the stages from data collection and preprocessing to model implementation and evaluation. Each step in the flowchart is clearly labeled to correspond with the detailed descriptions provided in the text, shown in figure 1.

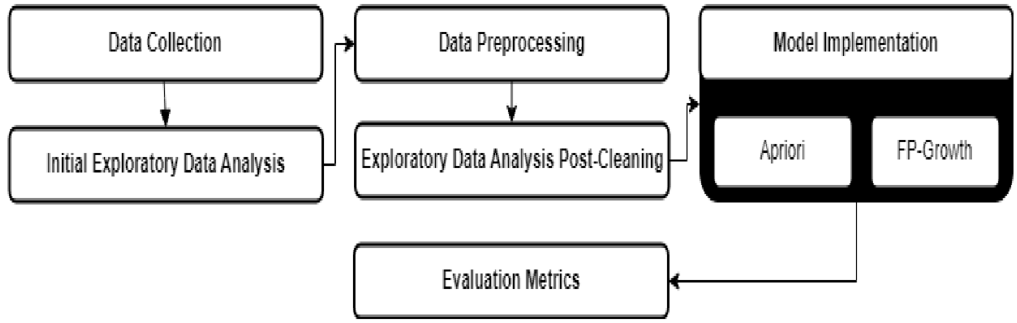

Figure 1 **Research Method**

The method begins with data collection, detailing how transaction data from "The Bread Basket" bakery is gathered. It proceeds to the initial exploratory data analysis (EDA), where the raw data is examined to identify any missing values, duplicates, and outliers. Following this, the data preprocessing stage involves cleaning and transforming the data to ensure it is ready for analysis. After preprocessing, a post-cleaning EDA is conducted to verify the effectiveness of the cleaning steps. The next stage involves the implementation of the Apriori and FP-Growth algorithms, where the models are trained and evaluated using specified metrics.

Data Collection

The dataset used in this study originates from The Bread Basket, a bakery located in Edinburgh. It encompasses customer transactions recorded from January 26, 2011, to December 27, 2018. The dataset was collected and provided with basic preprocessing to ensure data consistency. It contains 20,507 entries, representing over 9,000 unique transactions. Key features include TransactionNo, Items, DateTime, Daypart, and DayType.

The dataset consists of several important variables. The TransactionNo serves as a unique identifier for each transaction, ensuring that each entry can be distinctly recognized. The Items variable lists the products purchased in each transaction, providing insight into customer preferences and purchasing patterns. The DateTime variable records the exact timestamp of each transaction, allowing for detailed temporal analysis of sales data. The Daypart variable indicates the part of the day when the transaction occurred, categorized as morning, afternoon, evening, or night. This helps in understanding the distribution of transactions throughout the day. The DayType variable specifies whether the transaction occurred on a weekend or a weekday, which can be crucial for analyzing patterns related to customer behavior on different days of the week.

An initial examination of the dataset reveals the first few rows, showcasing transactions such as Bread, Scandinavian, Hot chocolate, and Jam, all occurring in the morning on weekends. The dataset contains 20,507 entries, with 9,465 unique transactions, ensuring a rich dataset for analysis. Summary statistics provide an overview of the dataset, indicating a range of transaction numbers from 1 to 9,684. The descriptive statistics are as follows: the transactions count is 20,507, with a mean transaction number of approximately 4,976. The standard deviation is about 2,796, indicating the variability in transaction numbers. The minimum transaction number is 1, the 25th percentile is at 2,552, the median (50th percentile) is 5,137, the 75th percentile is 7,357, and the maximum transaction number is 9,684. These statistics offer a comprehensive snapshot of the transaction distribution within the dataset.

Examining the unique values in the 'Items' column reveals a diverse array of products sold at the bakery, including Bread, Scandinavian, Hot chocolate, Jam, Cookies, and many others, indicating a broad product range. The dataset spans from January 11, 2016, to December 3, 2017. The unique values in the 'Daypart' column are Morning, Afternoon, Evening, and Night, with most transactions occurring in the morning and afternoon. The 'DayType' column categorizes transactions into 'Weekend' and 'Weekday', showing a higher frequency of weekday transactions. Specifically, the value counts for 'Daypart' are: Afternoon with 11,569 transactions, Morning with 8,404 transactions, Evening with 520 transactions, and Night with 14 transactions. For 'DayType', the counts are 12,807 transactions on weekdays and 7,700 on weekends. This distribution suggests that most customer activity is concentrated during the daytime on weekdays, which could be insightful for planning optimal marketing strategies and inventory management.

Initial Exploratory Data Analysis (EDA)

The initial EDA focused on examining the raw data to identify missing values, duplicates, and outliers. This step was crucial to understand the quality and characteristics of the dataset before proceeding with further analysis. The analysis revealed no missing values in any of the columns, indicating that the data was complete and did not require imputation. However, there were 1,620 duplicate rows, which needed to be addressed to ensure the accuracy of the analysis.

To gain a comprehensive understanding of the dataset, summary statistics were calculated. These statistics provided an overview of the data, including the distribution of daily transaction counts and the identification of the most frequently purchased items. The top ten most frequently purchased items were Coffee, Bread, Tea, Cake, Pastry, Sandwich, Medialuna, Hot chocolate, Cookies, and Brownie. Coffee was the most purchased item with 5,471 occurrences, followed by Bread with 3,325 occurrences. These frequent items were identified as potential outliers due to their high purchase frequency.

Descriptive statistics offered further insights into the dataset. The dataset contained 20,507 transactions with a mean transaction number of approximately 4,976. The transactions spanned from January 11, 2016, to December 3, 2017. The 'Daypart' column revealed that most transactions occurred in the afternoon, with 11,569 entries, followed by the morning with 8,404 entries. The 'DayType' column indicated that weekdays had a higher frequency of transactions (12,807) compared to weekends (7,700). This information was crucial for understanding the temporal distribution of sales.

Visualization played a significant role in the initial EDA. Histograms were used to display the distribution of transaction counts per day, revealing that the number of transactions per day varied, with an average of approximately 60 transactions per day as shown in figure 2.

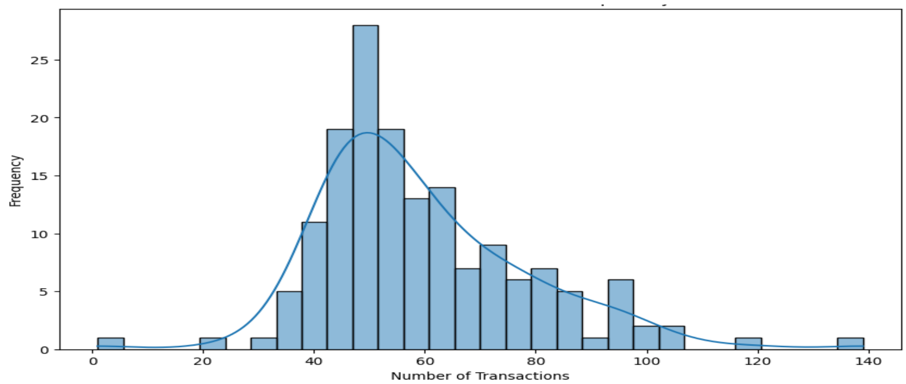

Figure 2 **Distribution of Transaction Counts per Day**

Bar plots illustrated the frequency of the top ten most purchased items, highlighting Coffee and Bread as the most popular products as shown in figure 3.

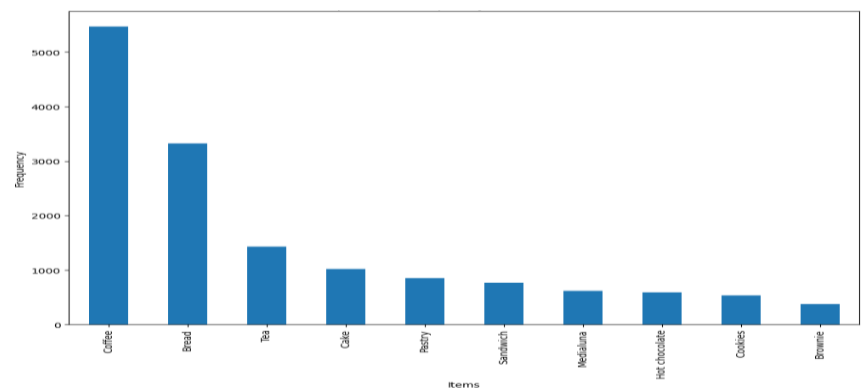

Figure 3 **Top 10 Most Frequently Purchased Items**

Data Preprocessing

In the data preprocessing stage, several steps were undertaken to prepare the dataset for analysis. The initial step involved data cleaning, where missing values were handled through imputation. However, since the dataset did not have any missing values in the relevant columns, imputation was not necessary. Duplicate transactions were identified and removed to ensure the integrity of the data. Additionally, item names were standardized by converting them to lowercase to maintain consistency across entries.

Following data cleaning, the text data in the 'Items' column underwent tokenization and stemming/lemmatization. Tokenization involved breaking down the text into individual tokens or words. The NLTK library was used to perform these tasks, and necessary packages such as 'punkt', 'wordnet', and 'stopwords' were downloaded. Stemming and lemmatization were then applied to reduce words to their base forms, which helps in standardizing the text and improving the accuracy of the subsequent analysis.

The final step in data preprocessing involved feature extraction using techniques like TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF was used to convert the processed text data into numerical features suitable for analysis. This method calculates the importance of a word in a document relative to its frequency across a collection of documents, providing a weighted representation of the text data. The resulting TF-IDF matrix was then converted into a DataFrame for easier manipulation and analysis.

The preprocessed data now consisted of 133 columns, each representing a unique item. The values in the DataFrame indicate the presence or absence of an item in each transaction. For example, the first five rows of the transformed data showed zeros for most items, indicating their absence in those transactions. This structured numerical representation of the data was crucial for the subsequent stages of analysis and modeling. The data preprocessing steps ensured that the dataset was clean, consistent, and in a suitable format for the application of machine learning algorithms.

Post-Cleaning EDA

The post-cleaning EDA was conducted to confirm the effectiveness of the data cleaning steps and to gain further insights into the dataset. After the cleaning process, the first few rows of the cleaned dataset were examined, revealing transactions such as Bread, Scandinavian, Hot chocolate, and Jam, all occurring in the morning on weekends. The examination showed that there were no missing values in any of the columns, confirming that the dataset was complete and did not require further imputation. Additionally, the number of duplicate rows was recorded as 1,620, indicating that the data cleaning process had effectively identified and managed duplicates.

Updated summary statistics provided a comprehensive overview of the cleaned dataset. The dataset contained 20,507 transactions, with Coffee being the most frequently purchased item (5,471 occurrences), followed by Bread. The 'Daypart' column revealed that the majority of transactions occurred in the afternoon (11,569 entries), followed by the morning (8,404 entries). The 'DayType' column showed that weekdays had a higher frequency of transactions (12,807) compared to weekends (7,700).

The detailed statistics of the cleaned dataset are as follows: there were 20,507 entries for transaction numbers, items, and timestamps. The 'Items' column had 94 unique values, with Coffee being the most common item purchased 5,471 times. The 'Daypart' column had four unique values, with the afternoon being the most frequent at 11,569 occurrences, followed by the morning at 8,404 occurrences. The dataset's timestamps ranged from January 11, 2016, to December 3, 2017, with various other descriptive statistics indicating the distribution of transaction numbers. The mean transaction number was approximately 4,976, with a standard deviation of about 2,796. The minimum transaction number was 1, and the maximum was 9,684.

For the 'DayType' column, there were two unique values, with weekdays being the most frequent at 12,807 transactions, compared to weekends with 7,700 transactions. These statistics highlight the significant patterns and trends within the dataset, offering valuable insights into customer purchasing behavior and the temporal distribution of transactions.

The distribution of transaction counts per day was also examined, providing insights into the daily transaction volumes. The analysis revealed that the average number of transactions per day was approximately 60, with a standard deviation of 19, indicating some variability in daily transaction volumes. The minimum number of transactions in a single day was 1, while the maximum number reached 139. The detailed statistics of the daily transaction counts are as follows: the dataset included 159 days of transaction records. The 25th percentile was 47 transactions per day, the median (50th percentile) was 56 transactions per day, and the 75th percentile was 69.5 transactions per day. These statistics offer a comprehensive view of the daily transaction distribution, highlighting both typical daily volumes and the range of fluctuations that occur.

Visualizations were employed to illustrate changes in data distributions and relationships after the cleaning process. Histograms were used to display the distribution of transaction counts per day, highlighting the frequency and range of transactions over the observed period. Bar plots depicted the frequency of the top purchased items, emphasizing the dominance of Coffee and Bread in sales. Scatter plots provided a visual representation of the number of transactions over time, allowing for an assessment of trends and patterns.

Model Implementation

In this study, we compared the Apriori and FP-Growth algorithms for market basket analysis. The selection of these algorithms was based on their widespread use and effectiveness in identifying frequent itemsets and generating association rules. To optimize the performance of each algorithm, hyperparameters were carefully tuned. Specifically, the minimum support threshold was set to 0.01 for both algorithms, ensuring a balance between computational efficiency and the relevance of the generated rules.

The dataset was divided into training and testing sets using an 80/20 split ratio. This division allowed us to evaluate the performance of the algorithms on unseen data, enhancing the robustness of our findings. Additionally, cross-validation techniques were applied to further validate the results and ensure that the models were not overfitting.

The model training process involved using appropriate libraries and tools, such as scikit-learn and mlxtend for Python. The data was prepared by transforming it into a basket format, where each item was represented as a binary variable indicating its presence or absence in a transaction. This transformation was achieved using the `groupby`, `unstack`, and `applymap` functions in pandas. However, we encountered a deprecation warning suggesting that the `applymap` function with non-boolean types might result in worse computational performance and could be discontinued in the future. This issue was noted, and future implementations will consider using a DataFrame with boolean types.

For the Apriori algorithm, the frequent itemsets and association rules were generated using the `apriori` and `association_rules` functions from the mlxtend library. The training set results showed that items like Coffee frequently appeared with other items such as Alfajores, Brownie, Cake, Cookies, and Juice. Key metrics such as support, confidence, and lift were calculated for each rule. For instance, the rule {Alfajores} → {Coffee} had a support of 0.018885, confidence of 0.520000, and lift of 1.087090, indicating a moderate association between these items.

Similarly, the FP-Growth algorithm was applied to the same training set, and the frequent itemsets and association rules were generated. The results were comparable to those of the Apriori algorithm, with items like Coffee frequently co-occurring with Scone, Brownie, Sandwich, Medialuna, and Pastry. For example, the rule {Scone} → {Coffee} had a support of 0.017829, confidence of 0.519231, and lift of 1.085482.

The performance of both algorithms was evaluated on the testing set to assess their generalizability. The Apriori algorithm's testing set results included rules such as {Alfajores} → {Coffee} with a support of 0.022715, confidence of 0.623188, and lift of 1.302092. The FP-Growth algorithm produced similar results, with rules like {Truffles} → {Coffee} showing a support of 0.011094, confidence of 0.512195, and lift of 1.070183.

Throughout the process, we encountered repeated deprecation warnings from the mlxtend library, indicating that using DataFrames with non-boolean types might result in worse computational performance. This warning highlights the importance of ensuring that data types are appropriately managed to optimize computational efficiency.

In conclusion, the model implementation phase demonstrated that both the Apriori and FP-Growth algorithms are effective for market basket analysis. The results from the training and testing sets provided valuable insights into customer purchasing behavior, and the comparison between the two algorithms highlighted their respective strengths in terms of support, confidence, and lift. Future work will address the deprecation warnings by adopting recommended practices for data type management.

Evaluation Metrics

The performance of the Apriori and FP-Growth algorithms was evaluated using several key metrics, including execution time, memory usage, support, confidence, and lift. These metrics provided a comprehensive assessment of the algorithms' effectiveness and efficiency in generating meaningful association rules from the market basket data.

Execution time and memory usage were critical performance metrics that highlighted the computational efficiency of each algorithm. The execution time for the Apriori algorithm was 4.08 seconds, with a memory usage of 45.36 MiB. The FP-Growth algorithm had a slightly longer execution time of 4.15 seconds but demonstrated significantly lower memory usage at only 0.08 MiB. These results indicate that while both algorithms had comparable execution times, FP-Growth was much more efficient regarding memory usage.

Support, confidence, and lift were used to evaluate the quality of the association rules generated by each algorithm. Support measures the proportion of transactions that contain a particular itemset, confidence indicates the likelihood that a rule's consequent is purchased when its antecedent is purchased, and lift assesses the strength of the rule by comparing the observed support to the expected support if the antecedent and consequent were independent.

The Apriori algorithm produced rules such as {Alfajores} → {Coffee} with a support of 0.018885, confidence of 0.520000, and lift of 1.087090, indicating a moderate association between these items. Other notable rules included {Brownie} → {Coffee}, {Cake} → {Coffee}, and {Juice} → {Coffee}, each with varying levels of support, confidence, and lift. The highest confidence observed among these rules was 0.548276 for {Juice} → {Coffee}, with a corresponding lift of 1.146202, suggesting a relatively strong association. The relation between

Support vs Confidence for Apriori Algorithm shown in figure 4.

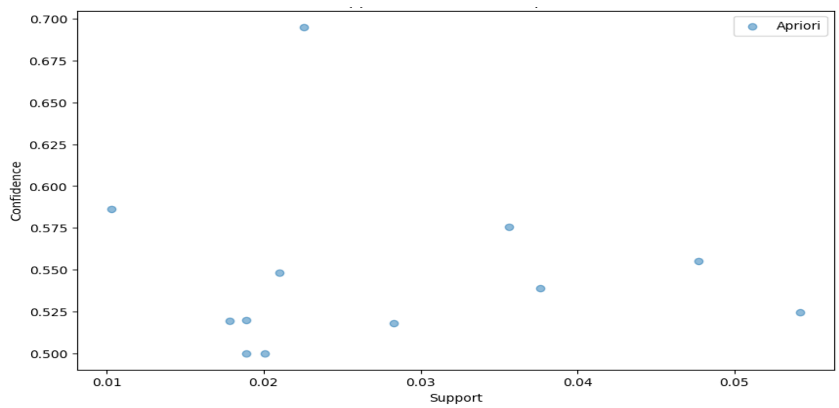

Figure 4 Support vs Confidence (Apriori)

Similarly, the FP-Growth algorithm generated comparable rules, such as {Scone} → {Coffee} with a support of 0.017829, confidence of 0.519231, and lift of 1.085482. Other significant rules included {Sandwich} → {Coffee}, {Medialuna} → {Coffee}, and {Pastry} → {Coffee}. The rule {Medialuna} → {Coffee} exhibited the highest confidence at 0.575693 and a lift of 1.203519, indicating a strong relationship between these items. The relation between Support vs Confidence for FP-Growth Algorithm shown in figure 5.

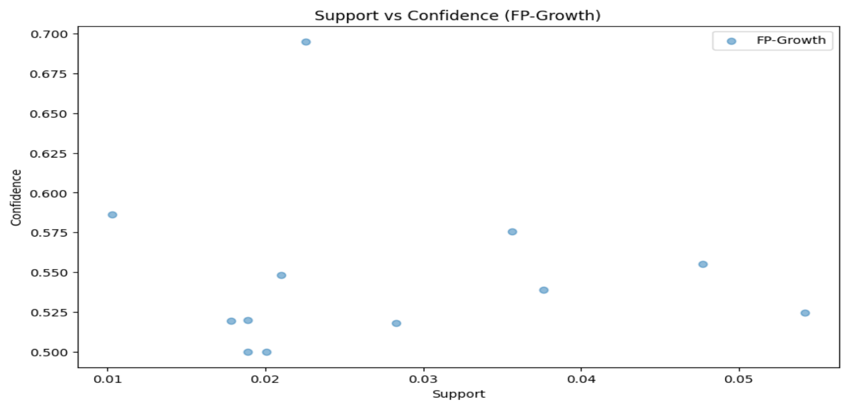

Figure 5 Support vs Confidence (FP-Growth)

Both algorithms displayed strengths in different areas. While Apriori was slightly faster in execution, FP-Growth's superior memory efficiency made it a more practical choice for larger datasets. The generated rules from both algorithms were similar regarding support, confidence, and lift, demonstrating their effectiveness in uncovering meaningful associations within the dataset.

Result and Discussion

Result

The performance of the Apriori and FP-Growth algorithms was thoroughly evaluated using several key metrics, including execution time, memory usage, support, confidence, and lift. The results of this evaluation were presented in both tabular and graphical formats to provide a comprehensive overview of the algorithms' performance.

Performance metrics for each algorithm were compiled and presented to clearly compare their efficiency and effectiveness. Execution time and memory usage were critical indicators of computational efficiency. The Apriori algorithm had an execution time of 4.08 seconds and a memory usage of 45.36 MiB, while the FP-Growth algorithm exhibited an execution time of 4.15 seconds and a significantly lower memory usage of 0.08 MiB. These metrics were visually represented through bar charts, highlighting the differences in computational resources required by each algorithm.

The detailed results included specific association rules generated by each algorithm, along with their corresponding support, confidence, and lift values. For instance, the Apriori algorithm produced rules such as {Alfajores} → {Coffee} with a support of 0.018885, confidence of 0.520000, and lift of 1.087090. Other notable rules included {Brownie} → {Coffee}, {Cake} → {Coffee}, and {Juice} → {Coffee}, each with varying levels of support, confidence, and lift. Visualizations such as bar charts and scatter plots were used to illustrate these results, clearly depicting the relationships between different items. For example, scatter plots of support versus confidence and lift versus confidence were created for both algorithms to visualize the distribution and strength of the generated rules. The FP-Growth algorithm generated similar association rules, such as {Scone} → {Coffee} with a support of 0.017829, confidence of 0.519231, and lift of 1.085482. Additional rules included {Sandwich} → {Coffee}, {Medialuna} → {Coffee}, and {Pastry} → {Coffee}. The highest confidence observed among these rules was 0.575693 for {Medialuna} → {Coffee}, with a corresponding lift of 1.203519. These results were also visualized through various plots, demonstrating the effectiveness of FP-Growth in identifying strong associations within the dataset.

A comparative analysis determined which algorithm performed best overall and in specific areas. The analysis considered both the efficiency and the quality of the generated rules. The Apriori algorithm demonstrated slightly faster execution times, but the FP-Growth algorithm excelled in memory efficiency, making it a more practical choice for larger datasets. Both algorithms generated high-quality rules with similar support, confidence, and lift values, indicating their robustness in market basket analysis.

The comparative analysis was visually represented through bar charts and scatter plots, facilitating an intuitive understanding of the differences and similarities between the two algorithms. The bar charts comparing execution time and memory usage highlighted FP-Growth's significant advantage in memory efficiency. Scatter plots comparing support versus confidence and lift versus confidence for both algorithms showcased the quality of the rules generated, with both algorithms producing strong and reliable associations.

Discussion

The comparative analysis of the Apriori and FP-Growth algorithms revealed that while both algorithms were effective in generating meaningful association rules, FP-Growth generally outperformed Apriori regarding memory efficiency. The primary reason for this performance difference lies in the intrinsic design and implementation of the algorithms. FP-Growth uses a more compact data structure known as the FP-tree, which reduces the need for multiple database scans and minimizes memory usage. This characteristic is particularly advantageous when dealing with large datasets, as it allows for more efficient processing and faster rule generation. The Apriori algorithm, on the other hand, relies on generating candidate itemsets and scanning the database multiple times, leading to higher memory consumption and longer execution times. The dataset's characteristics, including the high frequency of certain items and many unique items, further highlighted these differences, with FP-Growth's ability to handle such data more efficiently.

The insights derived from the best-performing algorithm, FP-Growth, have significant practical implications for bakery sales at "The Bread Basket." By identifying strong associations between frequently purchased items, the bakery can optimize product placement to encourage complementary purchases. For example, suppose the analysis reveals that customers who buy Coffee are also likely to buy Cake or Medialuna. In that case, these items can be strategically placed near each other to increase the likelihood of combined sales. Additionally, the bakery can refine its marketing strategies by targeting promotions and discounts on items frequently appearing together in transactions. Understanding these associations also aids in better inventory management, ensuring that popular item combinations are well-stocked to meet customer demand and reduce the risk of stockouts.

Despite the valuable insights gained from this study, several limitations must be acknowledged. The dataset's quality plays a crucial role in the accuracy of the analysis. Any inaccuracies or biases in the data can lead to misleading results. For instance, missing data or unrecorded transactions could affect the frequency and confidence of the generated rules. The study also assumes that past purchasing behavior indicates future behavior, which may only sometimes hold true due to changing customer preferences or external factors influencing sales. Additionally, the analysis was conducted on a specific dataset from a single bakery, which may limit the generalizability of the findings to other retail contexts.

Future research should explore hybrid approaches that combine the strengths of both Apriori and FP-Growth algorithms to enhance performance and accuracy. Integrating real-time data analysis could provide more dynamic and responsive insights, allowing businesses to adapt quickly to changing customer behaviors and market conditions. Expanding the scope of the study to include data from multiple retail sectors can help generalize the findings and develop more comprehensive strategies for market basket analysis. Further investigation into advanced techniques, such as incorporating customer segmentation and personalized recommendations, can also enhance the practical applications of the insights gained from association rule mining. By addressing these directions, future research can build on the current study's foundation and provide more robust and actionable business recommendations.

Conclusion

This study compared the Apriori and FP-Growth algorithms for discovering association rules from bakery sales data. The FP-Growth algorithm outperformed Apriori in memory efficiency and was preferred for large datasets due to its compact data structure. The study recommends using FP-Growth insights to optimize product placement, create targeted promotions, and improve inventory management. Limitations include data quality, assumptions about future behavior, and generalizability. Future research should explore hybrid approaches, incorporate real-time data, expand the study to multiple retail sectors, investigate customer segmentation, and address algorithm constraints. The study's findings and recommendations provide a solid foundation for improving sales and customer satisfaction through market basket analysis.

Declarations

Author Contributions

Conceptualization: H. and A.E.W.; Methodology: A.E.W.; Software: H.; Validation: H. and A.E.W.; Formal Analysis: H. and A.E.W.; Investigation: H.; Resources: A.E.W.; Data Curation: A.E.W.; Writing Original Draft Preparation: H. and A.E.W.; Writing Review and Editing: A.E.W. and H.; Visualization: H.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Bhanap and Dr. S. Kawthekar, “Data Mining for Business Intelligence in Social Network: A survey,” IARJSET, vol. 2, no. 12, pp. 129–131, Dec. 2015, doi: 10.17148/IARJSET.2015.21223.

[2] P. Devda, S. Shah, and M. Vasavada, “ANALYTICAL CRM FOR GOOGLE EDGE - DATA MINING FRAMEWORK WITH REFERENCE TO PHARMACEUTICALS INDUSTRY IN INDIA,” Int. J. Manag. Public Policy Res., vol. 2, no. 1, pp. 40–61, Jan. 2023, doi: 10.55829/ijmpr.v2i1.108.

[3] D. Acharjya, Sreekumar, R. Gopalan, and M. Desai, “Customer Classification in Indian Retail Sector- A Comparative Analysis of Various Machine Learning Approaches,” Int. J. Oper. Quant. Manag., vol. 26, pp. 1–28, Mar. 2020, doi: 10.46970/2020.26.1.1.

[4] U. G. Çiçekli and İ. Kabasakal, “Market Basket Analysis of Basket Data With Demographics: A Case Study in E-Retailing,” Alphanumeric J., vol. 9, no. 1, pp. 1–12, 2021, doi: 10.17093/alphanumeric.752505.

[5] S.-H. Liao and Y.-S. Tasi, “Big Data Analysis on the Business Process and Management for the Store Layout and Bundling Sales,” Bus. Process Manag. J., 2019, doi: 10.1108/bpmj-01-2018-0027.

[6] U. G. Dastidar, S. S. Ambekar, M. Hudnurkar, and A. D. Lidbe, “Experiential Retailing Leveraged by Data Analytics,” Int. J. Bus. Intell. Res. IJBIR, vol. 12, no. 1, pp. 98–113, Jan. 2021, doi: 10.4018/IJBIR.20210101.oa6.

[7] E. Bilgiç, Ö. K. Çakir, M. Kantardzic, and Y. Duan, “Retail Analytics: Store Segmentation Using Rule-Based Purchasing Behavior Analysis,” Int. Rev. Retail Distrib. Consum. Res., vol. 31, no. 4, pp. 457–480, 2021, doi: 10.1080/09593969.2021.1915847.

[8] H. Wang and P. Budsaratragoon, “Exploration of an “Internet”; Grounded Approach for Establishing a Model for Evaluating Financial Management Risks in Enterprises”, Int. J. Appl. Inf. Manag., vol. 3, no. 3, pp. 109–117, Sep. 2023.

[9] I. Ismarmiaty and R. Rismayati, “Product Sales Promotion Recommendation Strategy with Purchase Pattern Analysis FP-Growth Algorithm,” Sink. J. Dan Penelit. Tek. Inform., vol. 7, no. 1, Art. no. 1, Jan. 2023, doi: 10.33395/sinkron.v8i1.11898.

[10] A. Patel and K. K. Tiwari, “Approach to Kirana Store Product Arrangement Using Machine Learning,” Asian J. Res. Comput. Sci., pp. 72–83, Sep. 2021, doi: 10.9734/ajrcos/2021/v11i430271.

[11] V. Sharma, B. Kedia, V. Yadav, and S. Mishra, “Tapping the Potential Space-Positioning of Private Labels,” J. Indian Bus. Res., vol. 12, no. 1, pp. 43–61, 2020, doi: 10.1108/jibr-09-2019-0263.

[12] Y. Shi, “Formulation and Implementation of a Bayesian Network-Based Model”, Int. J. Appl. Inf. Manag., vol. 3, no. 3, pp. 101–108, Sep. 2023.

[13] Research Scholar, Department of Computer and Information Science, Annamalai University, Chidambaram (Tamil Nadu), India, P. Yoganandhini, G. Prabakaran, and Associate Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram (Tamil Nadu), India, “Market Basket Analysis with Enhanced Support Vector Machine (ESVM) Classifier for Key Security in Organization,” Int. J. Eng. Adv. Technol., vol. 9, no. 2, pp. 3261–3267, Dec. 2019, doi: 10.35940/ijeat.B3186.129219.

[14] A. P. Wibawa et al., “Mean-Median Smoothing Backpropagation Neural Network to Forecast Unique Visitors Time Series of Electronic Journal,” J. Appl. Data Sci., vol. 4, no. 3, pp. 163–174, Aug. 2023

[15] M. Kaur and S. Kang, “Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining,” Procedia Comput. Sci., vol. 85, pp. 78–85, 2016, doi: 10.1016/j.procs.2016.05.180.

[16] M. Szymkowiak, T. Klimanek, and T. Józefowski, “Applying Market Basket Analysis to Official Statistical Data,” Econometrics, vol. 22, no. 1, pp. 39–57, 2018, doi: 10.15611/eada.2018.1.03.

[17] S. N. Maharani, B. Sugeng, M. Makaryanawati, and M. M. Ali, “Bank Soundness Level Prediction: ANFIS vs Deep Learning,” J. Appl. Data Sci., vol. 4, no. 3, pp. 175–189, Sep. 2023,

[18] S. Rana and Md. N. I. Mondal, “Seasonal and Multilevel Association Based Approach for Market Basket Analysis in Retail Supermarket,” Eur. J. Inf. Technol.

Comput. Sci., vol. 1, no. 4, pp. 9–15, 2021, doi: 10.24018/compute.2021.1.4.31.

[19] G. Lisanawati and J. E. Kehinde, “When Technology Meets Money Laundering, What Should Law Do? New Products and Payment Systems and Cross Border Courier,” Int. J. Informatics Inf. Syst., vol. 5, no. 3, pp. 142–149, Sep. 2022

[20] E. Canales, L. Fan, D. R. Buys, and M. D. Cantave, “A Market Basket Assessment: Prices and Availability of Healthy Foods Across SNAP-Authorized Food Outlets in Counties With High Obesity Rates in Mississippi,” Prev. Chronic. Dis., vol. 18, 2021, doi: 10.5888/pcd18.210173.