Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

FP-Growth Algorithm: Mining Association Rules without Candidate Sets Generation

Ma'aruf Mohammed Lawal[1] and Ogedengbe Tunde Matthew[2]
[1]Ahmadu Bello University, Zaria, [2]Joseph Sarwuan Tarka, University, Makurdi
mmlawal80@gmail.com[1], matt.oged@uam.edu.ng[2]

Abstract

Over the years, due to modern technological advancement, unprecedented volume of data is been captured, and this has necessitated the need to mine such data to provide decision-based solution to non-trivial problems. Deploying an efficiently critical decision-based solution for handling such problems, require data mining algorithms. These evolving techniques emerged as an indispensable tools for pattern discovery in inventory data. With one notable technique being the application of Association Rule analysis, especially the Market Basket Analysis. However, mining association rules from large datasets can be daunting due to the volume of candidate sets generated by association rule algorithms like Apriori and ECLAT. Thus, candidate sets generated by these association rule based algorithms yield numerous rules, which contain both interesting and uninteresting ones. Hence, making interpretation overwhelming and decision-making challenging. On this note, this paper focused on demonstrating the efficiency of the FP-Growth algorithm in extracting relevant and interesting association rules for mining transaction itemsets over large datasets. By examining the FP-Growth algorithm design, functionality, and performance in depth analysis. The FP-Growth algorithm, which is an improved version of the Apriori algorithm is introduced with the intent to reduce the overhead costs by employing the FP-Tree data structure that efficiently encode the frequency information of itemsets in a dataset. To demonstrate performance improvement of the FP-Growth over the Apriori algorithms, the two algorithm were implemented on the WEKA data mining platform using a supermarket dataset. The performance of both algorithms is evaluated and compared in terms of computational time. The experimental results shows that the FP-Growth algorithm recorded 82.04% improvement over the Apriori algorithm. This improvement is attributed to the FP-Growth algorithm single dataset scan and the absence of candidate set generation which is inherent in the Apriori algorithm.

Keywords: *FP-Growth*; *Itemset; Candidate-set; Confidence; Association Rules.*

1. INTRODUCTION

Over the decade, modern computing has been largely responsible for influencing the growth of data been generated. The quest for deploying solutions for handling critical decision-based problem efficiently over such high dimensional data, has resulted in the introduction of a number of data mining algorithms. Additionally, the advent of big data has highlighted the importance of efficient and scalable algorithms to facilitate data-driven decision-making across numerous domains, such as marketing, healthcare, and finance (Nasyuha et al., 2020). While traditional algorithms like *Apriori* algorithm have been widely used, they often face limitations in handling large-scale datasets with numerous transactions and items.

Despite the long use of Apriori algorithm to mine frequent itemsets and generate association rules (Pradana et al., 2022), there is significant performance bottlenecks and hindering associated with the applicability of Apriori algorithm in real-world scenarios. This approach is computationally expensive when datasets contain a large number of transactions and items. This is not surprising, as it generates candidate itemsets by utilizing a level-wise search strategy, and similarly prune the generated candidate itemsets iteratively based on their support counts (Alcan et al., 2022). To overcome this challenge,

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

researchers have proposed various alternative approaches, including the *FP-Growth algorithm* (Han et al., 2004), which has gained significant attention due to its innovative use of a compact data structure called the *FP-Tree* (Wu & Liu, 2019).

1.1 Background of the Study

Data mining algorithms emerge as indispensable tools for pattern discovery in inventory data, with one notable technique being the application of Association Rule analysis, specifically Market Basket Analysis (Hartomo *et al*., 2020; Sornalakshmi *et al*., 2021). Among data analysis methods, association rule techniques pose particular challenges, yet they wield significant potential in uncovering hidden patterns within large datasets. To extract valuable insights from data for informed business decision-making, data mining algorithms leveraged on statistical and mathematical techniques (Wongwan & Laosiritaworn, 2018; Salam *et al*., 2018; Agapito *et al*., 2019). Particularly, data mining process finds application across various data types, including customer data, inventory, transactions, and marketing data. This process encompasses several key steps, from data preprocessing to method selection, execution, result evaluation, and interpretation (Napitupulu *et al*., 2019; Fitriah *et al*., 2023).

Most of the proposed algorithms for data mining process employ diverse array of data mining techniques exists, including classification, regression, clustering, association, and ranking models, each tailored for distinct purposes and data types (Kasim & Riadi, 2017; Anas *et al*., 2022). In the business environment, data mining serves as a potent tool for optimization, enabling enterprises to streamline operations, enhance decision-making, and unlock untapped potential for growth and efficiency (Vijayarani & Sharmila, 2016).

Several research works for pattern discovery in inventory data employ the Market Basket Analysis. This type of data mining algorithm stands as a pivotal data mining technique employed to convert customer purchasing patterns, particularly in the context of supermarket shopping (Agrawal & Srikant, 1994; Zhang *et al*., 2019; Unvan, 2021; Idris *et al*., 2022). This analytical method focuses on identifying relationships among items present in a customer's shopping cart, aiming to unveil concurrent purchasing habits. In the current study, Market Basket Analysis is conducted utilizing the Frequent Pattern Growth (*FP-Growth*) algorithm (Dong *et al*., 2019; Bagui *et al*., 2020; Yudha *et al*., 2020) provided in association rule mining. Thus, enabling the extraction of meaningful patterns and relationships from large datasets (Agapito *et al*., 2019).

Association rule mining is a fundamental technique in data mining and knowledge discovery, enabling the extraction of meaningful patterns and relationships from large datasets (Agapito *et al*., 2019). Such insights play a crucial role in numerous domains, including retail, healthcare, finance, and more, where the understanding of hidden patterns can significantly impact decision-making and strategy development (Bagui *et al*., 2020). However, the effectiveness of association rule mining relies heavily on the efficiency and scalability of the underlying algorithms, particularly when dealing with the increasing volume and complexity of modern datasets (Hu *et al*., 2021).

To address these limitations, researchers have proposed various alternative algorithms that aim to improve efficiency and scalability in association rule mining. One of such algorithm is the FP-Growth (Frequent Pattern Growth) algorithm, which offers a novel approach to data mining by eliminating the need for generating candidate itemsets (Ghassani et al., 2021; Dwiputra et al., 2023; Putri & Hardianto, 2024). Instead, *FP-Growth* algorithm constructs a compact data structure known as the FP-Tree, which

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

efficiently represents the dataset; enables direct mining of frequent itemsets and generation of association rules (Han et al., 2000).

The data mining process involves generating a list of association rules comprising item combinations that meet predefined minimum support and confidence thresholds. The efficiency of *FP-Growth* algorithm lies in its ability to avoid challenges posed by large prospective itemsets, utilizing a tree-specific data structure known as the *FP-Tree*. This compact representation of data enables efficient organization and traversal, facilitating rapid identification of frequent itemsets (Ghassani et al., 2021; Komarasay et al., 2022). The *FP-Growth* algorithm approach mitigates potential issues arising from an excessively large number of prospective itemsets, ensuring streamlined and effective Market Basket Analysis. By leveraging on the *FP-Growth* algorithm capabilities, businesses can gain valuable insights into customer purchasing behaviour, informing strategic decisions related to product assortment, pricing strategies, and promotional activities (Hartomo et al., 2020). In essence, Market Basket Analysis powered by the *FP-Growth* algorithm empowers businesses to unravel intricate patterns within transactional data, enabling them to optimize operations, enhance customer satisfaction, and drive profitability in the competitive retail landscape (Erwin, 2009; Ardiantoro & Sunarmi, 2020).

1.2 Motivation

Despite the widespread adoption of *Apriori* algorithm or association rule mining, its limitations have become increasingly evident, particularly with the growing volume of modern datasets (Idris *et al*., 2022). *Apriori* algorithm relies on frequent database scans and the generation of numerous candidate sets, leading to higher computational time and memory usage (Thurachon & Kreesuradej, 2021).

The following hypothetical transaction dataset in **Table 1** is introduced to demonstrate the limitations of *Apriori* algorithm.

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Table 1: Transaction Data

Transaction ID	Items
T1	*E, K, M, N, O, Y*
T2	*D, E, K, N, O, Y*
T3	*A, E, K, M*
T4	*C, K, M, U, Y*
T5	*C, E, I, K, O*

The frequency of each item was determined by scanning the dataset to count the number of times each item occurs in each transaction as shown in **Table 2**;

Table 2: Items' Frequency

C1	Support (Frequency)
A	1
C	2
D	1
E	4
K	5
M	3
N	2
O	3
U	1
Y	3

Assuming a minimum support threshold of 3, then **Table 3** consists the list of frequent 1-itemset (*L1*) (i.e. those candidates that meet the minimum support threshold value)

Table 3: Frequent 1-itemset

L1	Support
E	4
K	5
M	3
O	3
Y	3

Next is the generation of candidate 2-itemset using lexicographic order, this is done by scanning the dataset to determine the support of each of the candidates generated as shown in **Table 4.**

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Table 4: Candidate 2-Itemset

C2	Support
EK	4
EM	2
EO	3
EY	2
KM	3
KO	3
KY	3
MO	1
MY	2
OY	2

Next is the generation of *L*2 itemset by pruning candidate 2-itemset, the candidates whose support is at least minimum support threshold value qualified for *L*2 as shown in **Table 5.**

Table 5: Frequent 2-Itemset

L2	Support
EK	4
EO	3
KM	3
KO	3
KY	3

Next is the generation of candidate 3-itemset with their corresponding support count from the dataset as depicted in **Table 6.** The dataset was scanned to determine where each of these item's combinations occur together, thus count the number of occurrences of the items.

Table 6: Candidate 3-Itemset

C3	Support
EKO	3
EKM	2
EKY	2
KMO	1
KMY	1
KOY	2

Only one Candidate 3-Itemset is frequent and qualified to generate *L*3 as shown in **Table 7.** The above I llustrations depict the limitation of Apriori algorithm.

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Table 7: Frequent 3-Itemset

L3	Support
EKO	3

Based on the working example typified in tables 1 to 7, the limitations of the Apriori algorithm are evident in its need for frequent database scans to generate candidate sets. This repeated scanning and the large number of candidates sets result in higher computational time and much memory is required when dealing a large dataset (Yudha *et al*., 2020). To this effect, there is the need to propose an efficient solution that requires low computational time in returning results over a large dataset. Association rule mining is a fundamental technique in data mining and knowledge discovery, enabling the extraction of meaningful patterns and relationships from large datasets (Agapito *et al*., 2019). Such insights play a crucial role in numerous domains, including retail, healthcare, finance, and more, where the understanding of hidden patterns can significantly impact decision-making and strategy development (Bagui *et al*., 2020). However, the effectiveness of association rule mining relies heavily on the efficiency and scalability of the underlying algorithms, particularly when dealing with the increasing volume and complexity of modern datasets (Hu *et al*., 2021).

Traditional association rule mining algorithms, such as Apriori, have long been used to mine frequent itemsets and generate association rules (Pradana *et al*., 2022). While traditional algorithms like *Apriori* have been widely used, they often face limitations in handling large-scale datasets with numerous transactions and items. To overcome these challenges, researchers have proposed various alternative approaches, including the *FP-Growth* algorithm (Han *et al*., 2004), which has gained significant attention due to its innovative use of a compact data structure called the *FP-Tree* (Wu & Liu, 2019).

The *FP-Tree*, as a prefix-tree structure, compresses the dataset by merging common prefixes of transactions, thus reducing the memory footprint and enabling more efficient processing (Putri & Hardianto, 2024). *FP-Growth* adopts a recursive divide-and-conquer strategy, identifying conditional patterns and constructing conditional *FP-Tree* for different items, thus enabling parallel processing and enhancing scalability (Fitriah *et al*., 2023). The *FP-Tree* facilitates efficient storage and processing of dataset information, enabling the direct mining of frequent itemsets and generation of association rules without resorting to candidate itemset generation (Ugwu & Udanor, 2021; Alcan *et al*., 2022). *FP-Growth* algorithm, renowned for its efficiency, is instrumental in extracting frequent itemsets from large datasets with minimal access to the original database. The algorithm operates based on two key parameters: support and confidence (Wei *et al.*, 2014; Yudha *et al*., 2020). The support parameter gauges the prevalence of item combinations within the dataset, while confidence measures the strength of association between individual items in the constructed association rules.

In addition to the development of the *FP-Growth* algorithm, researchers have also proposed extensions and improvements to address specific challenges and requirements in association rule mining. For example, the *TFP-Growth* algorithm enhances *FP-Growth* by incorporating top-down traversal and compression techniques, leading to faster mining performance on large datasets (Saputra *et al*., 2023). Another variant, the *FUP-Growth* algorithm, focuses on preserving user privacy in association rule mining by introducing a compact tree structure called the *FUP-Tree*, which allows for secure multi-party computation (Thurachon & Kreesuradej, 2021).

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Moreover, researchers have explored the integration of association rule mining with other data mining techniques and machine learning algorithms to enhance the mining process or the quality of extracted patterns (Dwiputra *et al*., 2023). For instance, the application of clustering algorithms prior to association rule mining can help identify groups of similar data instances, which may lead to the discovery of more relevant and interpretable patterns. Furthermore, ensemble methods that combine multiple association rule mining algorithms have been proposed to improve the overall performance and robustness of pattern extraction (Liu & Wu, 2018).

The primary objective of this research paper is to provide a comprehensive exploration of the *FP-Growth* algorithm and its effectiveness in mining association rules without the need for candidate sets. In general, the main contributions of this work are briefly describe as follows:

* To the best of our knowledge, no work has actually examined the *FP-Growth* algorithm design, functionality, and performance in depth analysis in order to determine the efficiency of *FP-Growth* algorithm with association rule mining over *Apriori* algorithm over large multidimensional dataset.

* We conduct a comparative analysis of the *FP-Growth* algorithm against the traditional candidate set-based algorithms (Apriori) in order to evaluate the performance and effectiveness of *FP-Growth* algorithm with association rule mining in real-world scenario.

* We aim to advance research work on association rule mining techniques by providing insights into the capabilities of the *FP-Growth* algorithm as a powerful tool for discovering meaningful patterns in large-scale datasets.

The rest of this paper is organized as follows: Related works, concept and fundamental theories of association rule mining, and analytical frameworks are presented in Section 2, while methodology employed in this research is presented in Section 3. The results and discussions is presented in Section 4, while Section 5 concludes this research work.

2 Related Works

This section presents the review of the related work in Subsection 2, while a working example is provided to show how *FP-Growth* algorithm generates its tree without resorting to candidate set generation in Subsection 2.1, and Subsection 2.2 showcase the main steps of the *FP-Growth* algorithm.

Association rule mining is a fundamental task in data mining, aim to discover interesting relationships or patterns in large transactional databases. Traditional algorithms for association rule mining, such as Apriori, generate candidate itemsets by iteratively scanning the database. This can lead to significant high computational overhead cost, especially for datasets with a large number of transactions and items (Dwiputra *et al*., 2023; Putri & Hardianto, 2024).

The *FP-Growth* (Frequent Pattern Growth) algorithm, proposed by Han *et al*. (2000), introduced a revolutionary approach to association rule mining by eliminating the need for generating candidate itemset. Instead, *FP-Growth* employs an *FP-Tree* data structure to efficiently encode the frequency information of itemsets in the dataset (Saputra *et al*., 2023; Zhang *et al*., 2019). By employing a divide-and-conquer strategy, the *FP-Growth* algorithm recursively mines frequent itemsets directly from the *FP-Tree*. Thus, significantly reducing the computational complexity compared to the *Apriori* algorithm.

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Several studies have highlighted the advantages of the *FP-Growth* algorithm over *Apriori* algorithm, in terms of computational efficiency, scalability, and memory usage (Borgelt 2005, Lin *et al*., 2011, Liu *et al*. 2018, Shawkat *et al*., 2022, Zeng *et al*., 2015, Zhang *et al*., 2019). Agrawal and Srikant (1994) first introduced the concept of association rule mining and proposed the *Apriori* algorithm, which utilizes a level-wise approach to generate candidate itemsets. However, the *Apriori* algorithm suffers from limitations such as multiple database scans and the generation of a large number of candidate itemsets (Shawkat *et al*., 2022; Borgelt, 2005).

Lin *et al*., (2011), introduced the *IFP-Growth* algorithm, an enhanced version of the *FP-Growth* algorithm. This work aim to improve the performance in association rule mining. The *IFP-Growth* algorithm utilizes an address-table structure and *FP-tree+* to reduce complexity and memory requirements. Thus, outperforming the *FP-Growth* algorithm in terms of memory space and execution time. Experimental results show that the *IFP-Growth* algorithm is efficient and suitable for high-performance applications in association rule mining, as it requires significantly less memory and exhibits fast execution time.

Furthermore, the *FP-Growth* algorithm have been successfully applied in diverse domains, including Market Basket Analysis, bioinformatics, web mining, and network traffic analysis. For example, a study by Liu *et al*. (2018) utilized the *FP-Growth* algorithm to for analyze gene expression data, showcasing its efficiency in identifying significant gene patterns associated with specific biological conditions.

Nasyuha *et al*., (2020), in their work emphasizes the importance of product arrangement in stores to attract customers and boost sales. It introduces data mining methods, including the *FP-Growth* algorithm, as tools for optimizing product displays. The *FP-Growth* algorithm outperform the *Apriori* algorithm, as it efficiently identifies frequent itemsets in datasets, and also offers potential benefits for adjusting product layouts based on customer preferences. This apply *FP-Growth* algorithm to maximize the effectiveness of product arrangements in retail stores and enhance the shopping experience.

Wicaksono *et al*., (2020), focused on enhancing the efficiency of association rule mining, a data mining method aimed at discovering relationships between items, by incorporating preprocessing techniques into the *Apriori* algorithm. Association rule mining typically involves two phases: identifying frequent patterns that meet minimum frequency criteria and extracting strict rules based on minimum support and confidence thresholds. However, traditional methods suffer from high memory usage and time consumption. By integrating aggregate functions into the *Apriori* algorithm, the study was able to address high memory usage issue associated with *Apriori* algorithm.

Thurachon & Kreesuradej (2021), introduced a novel approach for incremental association rule mining called *FIUFP-Growth* algorithm. This algorithm utilizes an *Incremental Conditional Pattern tree* (*ICP-tree*) and a compact sub-tree to efficiently retrieve previous frequent itemsets from the original database. Thus, the algorithm minimizes unnecessary rescans, reducing resource usage and execution time compared to existing methods. Experimental results depict a 46% improvement in execution time over the *FP-Growth*, *FUFP-tree*, *Pre-FUFP*, and the *FCFPIM* algorithms at a 3% minimum support threshold. This approach offers significant benefits for computer business intelligence methods by facilitating faster and more efficient incremental association rule mining.

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Ugwu & udanor (2021), addressed the critical role of Customer Relationship Management (CRM) in enhancing customer engagement with an increased business profit. It highlighted the importance of having an effective collaboration and interaction between organizations and customers in identifying customer needs while maintaining trust and loyalty. To address these challenges, the study employed an association rule learning that utilizes the frequent pattern growth algorithm to identify patterns in customer purchasing behavior. The *Object-Oriented Analysis and Design methodology* (*OOAD*) is utilized for system analysis and design, while implementation is carried out using Python and MySQL. The contribution of the work lies in enabling firms to understand customer interests and preferences, thereby facilitating customer retention and fostering long-term business relationships.

Komarasay *et al.*, (2022), addressed the challenges posed by the complex nature of data generated daily through electronic gadgets. Their work emphasizes the need for an efficient big data analytics techniques using a supermarket dataset as a case study. Specifically, it focuses on the task of finding frequent pattern matching in large datasets. Similar works that address same issue often lack accuracy and suffer from long retrieval times. This is not surprising, as they do not preprocess the dataset. To address these limitations, the proposed work introduces an *Integrated Frequent Pattern* (*IFP*) Growth algorithm, which preprocesses the dataset and leverages on parallel processing on a Hadoop platform. By utilizing the *Hadoop Distributed File System* (*HDFS*) and multiprocessor capabilities, the *IFP* Growth algorithm aims was able to improve accuracy while minimizing prediction time.

2.1 FP-Growth Working Example

A working example is provide to show how *FP-Growth* algorithm generates its tree without resorting to candidate set generation. The Frequent Pattern set constructed includes elements with frequencies equal to or greater than the minimum support. These items are then organized in descending order according to their respective frequencies from **Table 3**. The resultant set L is: L = {*K*: 5, *E*: 4, *M*: 3, *O*: 3, *Y*: 3}

The Ordered-Item set for each transaction is constructed by iterating through the Frequent Pattern set and determining if the current item is present in the transaction as shown in **Table 8**. If the item is found, it is added to the Ordered-Item set for that transaction. This process results in the creation of the following table for all transactions:

Table 8: Ordered Itemsets

Transaction ID	Items	Ordered Itemset
T1	*E, K, M, N, O, Y*	*K, E, M, O, Y*
T2	*D, E, K, N, O, Y*	*K, E, O, Y*
T3	*A, E, K, M*	*K, E, M*
T4	*C, K, M, U, Y*	*K, M, Y*
T5	*C, E, I, K, O*	*K, E, O*

a) Insert the set {*K, E, M, O, Y*} into Tree data structure; Here, all items are sequentially linked together in the order of their occurrence in the set, and the support count for each item is initialized as 1, as shown in **Fig. 1.**

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

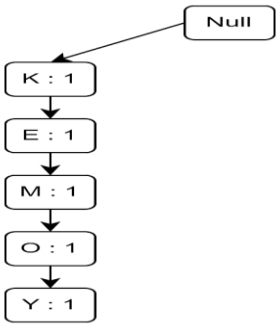

Fig. 1: First item linking

b) Inserting the set {*K*, *E*, *O*, *Y*}:
Up to the insertion of elements *K* and *E*, the support count is incremented by 1 as shown in **Fig. 2**. When inserting *O*, as there is no direct link between *E* and *O*, a new node is initialized for item *O* with a support count of 1, and item *E* is linked to this new node. Similarly, upon inserting *Y*, a new node is initialized for item *Y* with a support count of 1, and the new node of *O* is linked with the new node of *Y*.

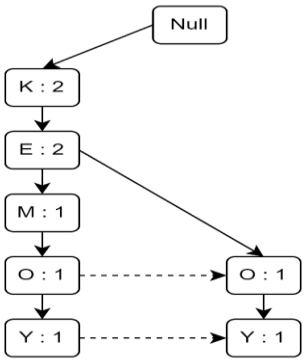

Fig. 2: Incrementing the count

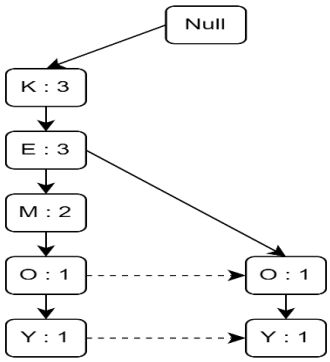

Fig. 3: Increasing support count

c) Inserting the set {*K*, *E*, *M*}: Here simply the support count of each element is increased by 1 in **Fig. 3**.

d) Inserting the set {*K*, *M*, *Y*}: Similar to step b), initially, the support count of *K* is incremented as shown in **Fig. 4**. Then, new nodes for *M* and *Y* are initialized, and they are linked accordingly.

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

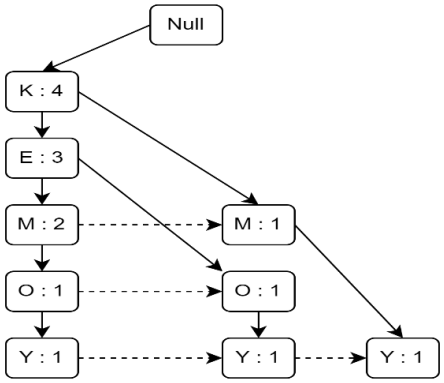

Fig. 4: More support counts

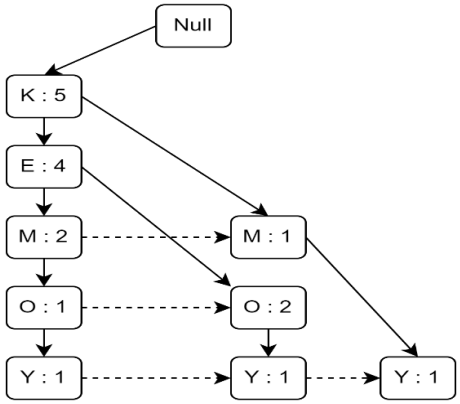

Fig. 5: Adding more nodes

e) Inserting the set {*K*, *E*, *O*}: Here, the support counts of the respective elements are simply incremented in **Fig. 5**. It's important to note that the support count of the new node for item *O* is also increased.

For each item, the Conditional Pattern Base is computed in **Table 9**, consisting of the path labels for all paths leading to any node of the given item in the frequent-pattern tree. It's worth mentioning that the items in the table below are arranged in ascending order of their frequencies.

Table 9: Conditional Frequent Pattern

Items	**Conditional Pattern Base**	**Conditional Frequent Pattern Tree**
Y	{*K*, *E*, *M*, *O*: 1}, {*K*, *E*, *O*:1}, {*K*, *M*:1}	{*K*:3}
O	{*K*, *E*, *M*: 1}, {*K*, *E*: 2}	{*K*, *E*: 3}
M	{*K*, *E*: 2}, {*K*:1}	{*K*:3}
E	{*K*: 4}	{*K*:4}
K		

From the Conditional Frequent Pattern tree, the Frequent Pattern rules are generated in **Table 10** by pairing the items of the Conditional Frequent Pattern Tree set with the corresponding item, as depicted in the table below.

Table 10: Frequent Pattern

Items	**Frequent Pattern Generated**
Y	{*K*, *Y*: 3}
O	{(K, O: 3), (*E*, *O*: 3), (*E*, *K*, *O*: 3)}
M	{(*K*. *M*: 3)}
E	{(*E*, *K*: 4)}
K	

For each row, two types of association rules can be inferred. For example, in the first row containing the elements *K* and *Y*, the rules $K \rightarrow Y$ and $Y \rightarrow K$ can be inferred. To determine the valid rule, the confidence of both rules is calculated, and the one with confidence greater than or equal to the minimum confidence value is retained.

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

2.2 Frequent Pattern-Growth Algorithm

The pseudocode outlines in Algorithm 1 are the main steps of the *FP-Growth* algorithm, which includes the construction of the *FP-Tree*, mining frequent itemsets recursively, generating conditional pattern bases, and constructing conditional *FP-Trees*. The algorithm efficiently identifies frequent itemsets in a dataset without generating candidate itemsets, making it suitable for association rule mining tasks. (Wei *et al.*, 2014; Putri & Hardianto, 2024).

Algorithm 1: *FP-Growth* Algorithm
Input: dataset D, min_support
Output: FP-Tree

```
FP-Growth (dataset D, minSupport) {
   // Construct the initial FP-Tree
   Initialize an empty FP-Tree
   for each transaction T in dataset D {
      Sort items in T by their support count in descending order
      Add T to the FP-Tree with proper support counts
   }

   // Mine frequent itemsets recursively
   Mine_Frequent_Itemsets(Empty_Set, FP-Tree, minSupport)
}
Mine_Frequent_Itemsets (Prefix, FP-Tree, minSupport) {

// Generate frequent itemsets from the FP-Tree
   for each item I in the header table of FP-Tree {
      if (Prefix ∪ {I} is frequent) {
         Output Prefix ∪ {I} as a frequent itemset
         Generate Conditional Pattern Base for I
         Construct Conditional FP-Tree from Conditional Pattern Base
         if (Conditional FP-Tree is not empty) {
            Mine_Frequent_Itemsets(Prefix ∪ {I},Conditional FP-Tree,minSupport)
         }
      }
   }
}
Generate_Conditional_Pattern_Base (item I, FP-Tree) {

   // Construct Conditional Pattern Base for item I
   Conditional Pattern Base = {}
   for each path P in FP-Tree where I appears {
      Extract suffix S from path P after item I
      Add S to Conditional Pattern Base with the support count of path P
   }
   return Conditional Pattern Base
}
Construct_Conditional_FP-Tree (Conditional Pattern Base) {

 // Construct Conditional FP-Tree from Conditional Pattern Base
   Conditional FP-Tree = Initialize an empty FP-Tree
   for each conditional pattern P in Conditional Pattern Base {
      Add P to Conditional FP-Tree with proper support counts
   }
   return Conditional FP-Tree
}
```

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

3. METHODOLOGY

This section provides details about the chosen algorithm, modifications, or adaptations made to accommodate the specific objectives of our study.

3.1 Dataset Selection

In this study, the selection of an appropriate dataset is crucial for conducting a comprehensive empirical evaluation of the *FP-Growth* algorithm. To this end, a supermarket transaction dataset from the Kaggle data mining repository was chosen as the primary dataset for analysis. The decision to utilize a supermarket transaction dataset is strategic, as it reflects a common and relevant application scenario for association rule mining techniques like *FP-Growth* algorithm. Supermarket transactions are characterized by the frequent purchase of multiple items in a single transaction, making them well-suited for Market Basket Analysis. The selected dataset comprises of 4627 instances and encompasses 217 attributes, representing a rich and diverse set of transactional data. These attributes are likely to encompass various aspects of customer transactions, including product identifiers, quantities purchased, transaction timestamps, and possibly additional metadata such as customer demographics or transaction types.

3.2 Experimental Setup

In this study, the experimental setup was designed to ensure the robustness and reproducibility of the empirical evaluation of the *FP-Growth* algorithm. For this purpose, WEKA (Waikato Environment for Knowledge Analysis) was adopted as the primary tool for experimentation and data analysis. WEKA is a renowned collection of machine learning and data analysis software, available free of charge under the GNU General Public License. Its user-friendly interface, extensive functionality, and robust performance make it a popular choice for researchers and practitioners alike.

In the experimentation phase, specific parameters were configured to execute the *FP-Growth* algorithm within the WEKA environment. These parameters include:

i. Minimum Support: A minimum support threshold of 0.1 was set to filter out infrequent itemsets and focus on identifying patterns that occur frequently within the dataset. This threshold ensures that only significant associations are considered during the mining process.

ii. Minimum Confidence: A minimum confidence threshold of 0.75 was established to determine the strength of association between items in the generated rules. This threshold indicates the minimum level of confidence required for a rule to be considered actionable and reliable.

iii. Number of Top Rules: The expected number of rules to be generated was set to 30. By limiting the number of rules to the top 30, the analysis focuses on the most relevant and informative patterns, facilitating easier interpretation and decision-making.

By configuring these parameters, the experimental setup aims to strike a balance between computational efficiency and the quality of generated association rules. These parameters are chosen based on domain knowledge, prior research, and empirical insights to ensure meaningful and actionable results. The adoption of WEKA as the experimentation platform, coupled with carefully chosen parameter settings,

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

lays the groundwork for a rigorous and insightful analysis of the *FP-Growth* algorithm performance and effectiveness in association rule mining tasks within the context of the selected supermarket transaction dataset.

3.3 Performance Metrics

Association rule mining is a fundamental data mining method employed to extract significant associations or correlations between itemsets within extensive datasets. For formal representation, let D denote a dataset encompassing all elements in the database. Each attribute of a record in D is designated as an item, collectively forming an itemset. Transactions, denoted as T, represent nonempty records (Liu & Wu, 2018). In a transaction T, consider two items, X and Y, both proper subsets of T. If X and Y are nonempty subsets with an empty intersection, the association rule $X \rightarrow Y$ is established within the item set T.

In this study, the performance of the *FP-Growth* algorithm is evaluated using a set of comprehensive metrics to assess its effectiveness in association rule mining. The chosen performance metrics encompass key aspects of rule quality, relevance, and significance, providing valuable insights into the algorithm's performance. The following metrics are utilized in evaluating the performance of *FP-Growth* algorithm against *Apriori* algorithm.

3.3.1 Support

This is a vital metric used to quantify the frequency of occurrence of a particular item in a dataset. Mathematically, the support of an item$X$, denoted as($Support(X)$), is calculated as the ratio of the number of transactions containing X to the total number of transactions in the dataset D as shown in Equation 1. It provides insights into the popularity of an item set and is a fundamental measure for identifying significant associations. Higher support values indicate a more prevalent occurrence of the item set, making it a key parameter in determining the relevance and strength of association rules (Wu & Liu, 2019).

$$Support(X) = \frac{Number\ of\ transactions\ contain\ item(X)}{Total\ \ number\ of\ transactions\ in\ D} \tag{1}$$

3.3.2 Confidence

This metric measures the reliability or strength of an association rule. Mathematically, the confidence of a rule $X \rightarrow Y$, denoted as $Confidence\ (X \rightarrow Y)$, is calculated as the ratio of the support of the item set $(X \cup Y)$ to the support of X as depicted in Equation 2.

$$Confidence(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)} \tag{2}$$

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

This formula quantifies the **likelihood** that if *X* occurs, *Y* will also occur in the same transaction. Confidence values range from 0 to 1, where a higher confidence indicates a stronger association between the item sets *X* and *Y*.

3.3.3 Lift

This is another metric that assesses the strength and significance of an association rule $X \rightarrow Y$ beyond what would be expected by chance. Mathematically, Equation 3 defines lift of a rule as the ratio of the confidence of the rule to the support of *Y*:

$$Lift(X \rightarrow Y) = \frac{Confidence(X \rightarrow Y)}{Support(Y)} \quad (3)$$

A *Lift* value greater than 1 suggests that the occurrence of *X* has a positive impact on the occurrence of *Y*, indicating a meaningful association. Conversely, a *Lift* less than 1 implies a weaker or potentially negative association. *Lift* provides valuable insights into the practical significance of association rules beyond individual support and confidence measures.

However, association rule mining involves two basic steps; generate frequent itemset from candidate set whose support value is at least a minimum support threshold and generate strong association rules from the frequent itemset based on minimum confidence threshold value (Shabtay *et al*., 2021).

4. RESULTS AND DISCUSSION

In this study, experiments were conducted on a supermarket transaction dataset comprising 4627 instances and 217 attributes. The performance of both the traditional *Apriori* algorithm and the *FP-Growth* algorithm was evaluated and compared in terms of computation time. **Fig. 6** illustrates the results of the experiments, highlighting the notable improvements in execution time achieved by the *FP-Growth* algorithm compared to *Apriori* across varying minimum support values. Specifically, *FP-Growth* algorithm demonstrates a significant percentage improvement of 82.04% in execution time over *Apriori* algorithm.

This performance *FP-Growth* algorithm over Apriori algorithm can be attributed to the elimination of candidate set generation, a process inherent in *Apriori* algorithm that necessitates multiple scans of the database to generate large candidate sets. By eliminating this step, *FP-Growth* algorithm minimizes the overhead associated computational cost with execution time, resulting in overall efficient performance. These findings stress the efficacy of the *FP-Growth* algorithm in association rule mining tasks, particularly in scenarios involving large-scale datasets and varying support thresholds. The reduced computation time recorded by *FP-Growth* algorithm makes it a compelling choice for real-world applications where efficiency and scalability are paramount. Conclusively, the experimental results validate the superiority of the *FP-Growth* algorithm over *Apriori* algorithm in terms of execution time. Thus, highlighting its potential to drive advancements in data mining and facilitate more efficient decision-making processes in diverse domain**s.**

Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 392-411
https://doi.org//10.47514/kjcs/2024.1.2.0016

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

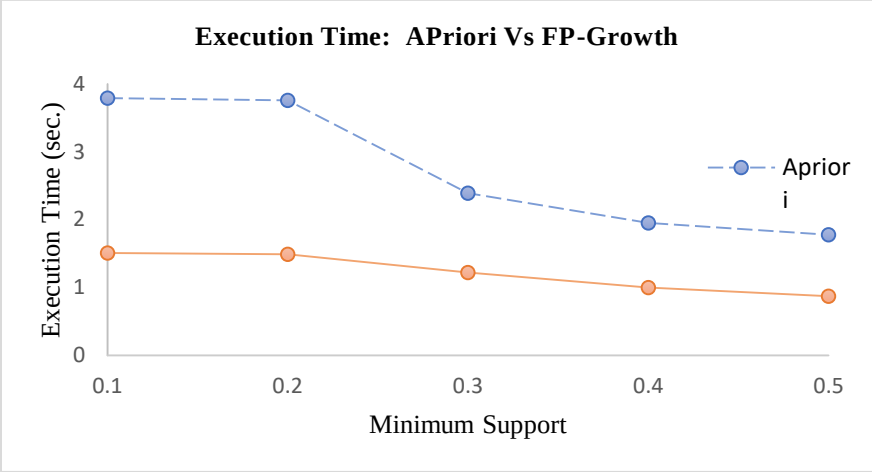

Fig. 6: Performance evaluation of *FP-Growth* Vs *Apriori* algorithm

5. CONCLUSION

In conclusion, this research investigated the performance and efficiency of the *FP-Growth* algorithm in association rule mining tasks, particularly in comparison to the traditional *Apriori* algorithm. Through empirical evaluation on a supermarket transaction dataset comprising 4627 instances and 217 attributes, several key insights were gleaned. The experimental results revealed significant improvements in execution time achieved by *FP-Growth* algorithm over Apriori, with *FP-Growth* algorithm demonstrating a remarkable percentage improvement of 82.04% across varying minimum support values. This improvement can be attributed to *FP-Growth* algorithm elimination of candidate set generation, thereby minimizing overhead and streamlining the mining process.

By leveraging on *FP-Growth* algorithm, businesses can extract meaningful patterns and insights from transactional data more efficiently, leading to informed decision-making and enhanced operational efficiency. The findings of this research accentuate the importance of algorithmic advancements in data mining techniques and their practical implications for real-world applications. The superiority of *FP-Growth* algorithm over *Apriori* algorithm in terms of execution time reaffirms its position as a cornerstone algorithm in association rule mining, paving the way for continued innovation and advancement in the field.

In future, further research could explore enhancements and optimizations of the *FP-Growth* algorithm, by investigating its applicability in different domains. Additionally, evaluates *FP-Growth* algorithm performance in conjunction with other data mining techniques. By continuously pushing the boundaries of algorithmic research, researchers and practitioners can unlock new opportunities that leverage on data-driven approaches for driving innovation and fostering growth across various sectors.

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Reference

Agapito, G., Milano, M., Guzzi, P. H. & Cannataro, M. (2019). "Mining Association Rules from Disease Ontology," Proceedings -2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, pp. 2239–2243.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (Vol. 1215, pp. 487-499).

Alcan, D., Ozdemir, K., Ozkan, B., Mucan, A. Y., & Ozcan, T. (2022). A comparative analysis of *Apriori* and *FP-Growth* algorithms for Market Basket Analysis using multi-level association rule mining. In Global Joint Conference on Industrial Engineering and Its Application Areas (pp. 128-137). Cham: Springer Nature Switzerland.

Anas, S., Rumui, N., Roy, A. & Saputro, P. H. (2022). "Comparison of *Apriori* Algorithm and *FP-Growth* in Managing Store Jurnal Pendidikan Teknologi Informasi, vol. 3, no. 1, pp. 118–129. [Online]. Available: http://journal.umkendari.ac.id/ index.php/decode/article/view/132

Ardiantoro, L., & Sunarmi, N. (2020). Badminton player scouting analysis using Frequent Pattern growth (*FP-Growth*) algorithm," in Journal of Physics: Conference Series, vol. 1456, no. 1. Institute of Physics Publishing.

Bagui, S., Devulapalli, K., & Coffey, J. (2020). "A heuristic approach for load balancing the *FP-Growth* algorithm on MapReduce," Array, vol. 7, p. 100035.

Borgelt, C. (2005). An Implementation of the *FP-Growth* Algorithm. In Proceedings of the 1st international workshop on open-source data mining: frequent pattern mining implementations (pp. 1-5).

Dong, S., Liu, M. & Zhang, S. (2019). "Association rules mining of silk relics database with the RC*FP-Growth* algorithm," Chinese Control Conference, CCC, vol. pp. 7804–7809.

Dwiputra, D., Widodo, A. M., Akbar, H., & Firmansyah, G. (2023). Evaluating the Performance of Association Rules in *Apriori* and *FP-Growth* Algorithms: Market Basket Analysis to Discover Rules of Item Combinations. Journal of World Science (JWS), 2(8), 1229-1248.

Erwin, E. (2009).Analisis Market Basket Dengan Algoritma *Apriori* dan *FP-Growth*," Generic, vol. 4, no. 2, pp. 26–30, [Online]. Available: http://generic.ilkom.unsri.ac.id/index.php/generic/article/view/15

Fitriah, I., Riadi, & Herman, (2023). "Analisis Data Mining Sistem Inventory Menggunakan Algoritma Apriori," Decode: Jurnal Pendidikan Teknologi Informasi, vol. 3, no. 1, pp. 118–129. [Online]. Available: http://journal.umkendari.ac.id/index.php/decode/article/view/132

Ghassani, F. Z., Jamaludin, A., & Irawan, A. S. Y. (2021). Market Basket Analysis Using the *FP-Growth* Algorithm to Determine Cross-Selling. Jurnal Informatika Polinema, 7(4), 49-54.

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. ACM sigmod record, 29(2), 1-12.

Hartomo, K. D., Yulianto, S., & Suharjo, R. A. (2020). Prediksi Stok dan Pengaturan Tata Letak Barang Menggunakan Kombinasi Algoritma Triple Exponential Smoothing dan *FP-Growth*," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 7, no. 5, pp. 869–878. https://www.ejurnal.stmik-budidarma.ac.id/index.php/ijics/article/view/4535

Hu, S., Liang, Q., Qian, H., Weng, J., Zhou, W. & Lin, P. (2021). Frequent-pattern growth algorithm-based association rule mining method of public transport travel stability," International Journal of Sustainable Transportation (IJST), vol. 15, no. 11, pp. 879–892, [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/15568318.2020.1827318

Idris, A. I., Sampetoding, E. A. M., Yoga, V., Ardhana, Maritsa, P.I., Sakri, A., Ruslan, H. & Manapa, E.S. (2022). "Comparison of Apriori, Apriori-TID and *FP-Growth* Algorithms in Market Basket Analysis at Grocery Stores," The International Journal of Informatics and Computer Science (IJICS), vol. 6, no. 2, pp. 107–112 [Online]. Available:

Kasim, H., & Riadi, I. (2017). "Detection of Cyberbullying on Social Media Using Data Mining Techniques," International Journal of Computer Science and Information Security (IJCSIS), vol. 15, pp. 244–250.

Komarasay, D., Gupta, M., Murugesan, M., Jenita Hermina, J., & Gokuldhev, M. (2022). Frequent Pattern Mining in Big Data Using Integrated Frequent Pattern (IFP) Growth Algorithm. In Proceedings of the 6th International Conference on Advance Computing and Intelligent Engineering: ICACIE 2021 (pp. 423-431). Singapore: Springer Nature Singapore.

Lin, K. C., Liao, I. E., & Chen, Z. S. (2011). An improved frequent pattern growth method for mining association rules. Expert Systems with Applications, 38(5), 5154-5161.

Liu, F. & Wu, G. (2018). "An improved *Apriori* algorithm based on compressed matrix," Journal of Shandong University (Engineering Edition), vol. 48, no. 6, pp. 82–88.

Napitupulu, G.T.M., Oktaviani, A., Sarkawi, D. & Yulianti, I. (2019). "Penerapan Data Mining Terhadap Penjualan Pipa Pada Cv.

Nasyuha, A. H., Jama, J., Abdullah, R., Syahra, Y., Azhar, Z., Hutagalung, J., & Hasugian, B. S. (2020). Frequent pattern growth algorithm for maximizing display items. Telkomnika (Telecommunication Computing Electronics and Control), 19(2), 390-396.

Pradana, M. R., Syafrullah, M., Irawan, H., Chandra, J. C., & Solichin, A. (2022). Market Basket Analysis Using *FP-Growth* Algorithm on Retail Sales Data. In 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) (pp. 86-89). IEEE.

Putri, I. D., & Hardianto, R. (2024). Application of the *FP-Growth* Algorithm in Consumer Purchasing Pattern Analysis. Journal of Computer Scine and Information Technology (JCSIT), 44-49.

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Salam, A., Zeniarja, J., Wicaksono, W. & Kharisma, L. (2018). "Pencarian Pola Asosiasi untuk Penataan Barang dengan Menggunakan Perbandingan Algoritma *Apriori* dan *FP-Growth* (Study Kasus Distro Epo Store Pemalang)," Jurnal DINAMIK, vol. 23, no. 2,

Saputra, J.P.B., Rahayu, S.A., & Hariguna, T. (2023). Market Basket Analysis using *FP-Growth* algorithm to design marketing strategy by determining consumer purchasing patterns. Journal of Applied Data Sciences, 4(1), 38-49.

Shabtay, L., Fournier-Viger, P., Yaari, R., & Dattner, I. (2021). A guided *FP-Growth* algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data. Information Sciences, 553, 353-375.

Shawkat, M., Badawi, M., El-ghamrawy, S., Arnous, R., & El-desoky, A. (2022). An optimized *FP-Growth* algorithm for discovery of association rules. The Journal of Supercomputing, 78(4), 5479-5506.

Sornalakshmi, M., Balamurali, S., Venkatesulu, M., Krishnan, M. N., Ramasamy, L. K., Kadry, S., & Lim, S. (2021) "An efficient *Apriori* algorithm for frequent pattern mining using mapreduce in healthcare data," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 1, pp. 390–403, [Online]. Available: https://beei.org/index.php/EEI /article/view /2096

Thurachon, W., & Kreesuradej, W. (2021). Incremental association rule mining with a fast incremental updating frequent pattern growth algorithm. IEEE Access, 9, 55726-55741. Transaction Data," International Journal of Computer and Information System (IJCIS), vol. 3, no. 4, pp. 158–162, [Online]. Available: http://www.ijcis.net/index.php/ijcis /article/view/96

Ugwu, N. V., & Udanor, C. N. (2021). Achieving effective customer relationship using frequent pattern-growth algorithm association rule learning technique. Nigerian Journal of Technology, 40(2), 329-339

Unvan, Y. A. (2021). "Market Basket Analysis with association rules," Communications in Statistics - Theory and Methods, vol. 50, no. 7, pp. 1615–1628, [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/03610926.2020. 1716255

Vijayarani, S., & Sharmila, S. (2016). "Comparative analysis of association rule mining algorithms," Proceedings of the International Conference on Inventive Computation Technologies, ICICT 2016

Wei, X., Ma, Y., Zhang, F., Liu, M., & Shen, W. (2014). Incremental *FP-Growth* mining strategy for dynamic threshold value and database based on MapReduce. In Proceedings of the 2014 IEEE 18th international conference on computer supported cooperative work in design (CSCWD) (pp. 271-276). IEEE.

Wicaksono, D., JAMBAK, M. I., & SAPUTRA, D. M. (2020). The comparison of *Apriori* algorithm with preprocessing and *FP-Growth* algorithm for finding frequent data pattern in association rule. In Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019) (pp. 315-319). Atlantis Press.

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Wongwan, K. & Laosiritaworn, W. (2018). "Application of association rules in woven wire mesh defects analysis," 2018 7th International Conference on Industrial Technology and Management, ICITM 2018, vol. 2018-January, pp. 325–329.

Wu, A. & Liu, A. (2019) "Improvement of *Apriori* algorithm based on Boolean matrix reduction," Computer engineering and science, vol. 9.

Yudha, T., Sunardi, S., & Fadlil, A. (2020). "Market Basket Analysis To Identify Stock Handling Patterns & Item Arrangement Patterns Using *Apriori* Algorithms," Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika, vol. 6, no. 1, pp. 33–41, [Online]. Available: https://journals.ums.ac.id/index.php/khif/article/view/8628

Zeng, Y., Yin, S., Liu, J., & Zhang, M. (2015). Research of improved *FP-Growth* algorithm in association rules mining. Scientific Programming, 2015, 6-6.

Zhang, G., Liu, C. & Men, T. (2019). "Research on data mining technology based on association rules algorithm," Proceedings of 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, ITAIC, pp. 526–530.