https://doi.org/10.22306/al.v11i2.504 *Received: 12 Nov. 2023; Revised: 27 Jan. 2024; Accepted: 06 Mar. 2024*

Determining consumer demand patterns for production planning using a data mining approach

A. Besse Riyani Indah
Universitas Hasanuddin, Faculty of Engineering, Department of Industrial Engineering, Jl. Malino No.8 F, Gowa, Sulawesi Selatan, Indonesia, a.besseriyani@unhas.ac.id (corresponding author)
Andi Siti Hajar Meliani
Universitas Hasanuddin, Faculty of Engineering, Department of Industrial Engineering, Jl. Malino No.8 F, Gowa, Sulawesi Selatan, Indonesia, andisitihajarmeliani@gmail.com
Saiful Mangngenre
Universitas Hasanuddin, Faculty of Engineering, Department of Industrial Engineering, Jl. Malino No.8 F, Gowa, Sulawesi Selatan, Indonesia, Saiful.ti@gmail.com
Andi Nurwahidah
Politeknik ATI Makassar, Departement of Agro-Industrial Engineering, Jl. Sunu No.220, Makassar, Sulawesi Selatan, Indonesia, Nurwahidah.andi@atim.ac.id
Arni Muarifah Amri
Institut Teknologi Telkom Surabaya, Faculty of Business and Information Technology, Department of Software Engineering, Jl. Ketintang 156, Gayungan, Surabaya, Jawa Timur, Indonesia, arnimuarifah@ittelkom-sby.ac.id

Keywords: data mining, clustering, K-Means Algorithm, association, FP-Growth Algorithm.

Abstract: The bread industry faces significant risks of losses in case of excess inventory. The initial stage in the K-Means Clustering Algorithm involves forming two clusters: C1 for slow-moving product data and C2 for fast-moving. Clustering products using the K-Means Algorithm resulted in Group 1 as slow-moving products with 44 types of items and Group 2 as fast-moving products with 15 types of items. It can be concluded that the bakery is experiencing losses due to an excess of overstocked products. After categorizing data into slow-moving and fast-moving groups, the subsequent phase involves employing the FP (Frequent Pattern)-Growth association rule algorithm to recognize consumer purchasing patterns. This algorithm aims to uncover relationships between items in a dataset and assess the probability of a person purchasing bread concurrently. By establishing a minimum support of 3% and a minimum confidence level of 30%, a total of 13 rules were generated, meeting the criteria for strong association rules. With this data, the store owner can specifically enhance inventory planning for fast-moving products by analyzing demand data and market trends. For slow-moving products, the store owner can adjust item placement or create product bundling with best seller items.

1 Introduction

The existence of business opportunities in the food industry means that micro, small and medium enterprises (MSMEs) are able to create independent sources of income. However, competition in the food industry is currently very tight, so business people are required to produce creative and innovative ideas, as well as manage existing resources so that they can achieve their main goals. One of the products from the food industry is bread which generally has characteristics that are easily damaged and have a relatively short expiration date so that the bread industry has a high risk of loss if there is excess stock.

From the findings of the assessments, it is evident that having surplus inventory poses a disadvantage for the company. This excess inventory is due to overproduction. Overproduction is the activity of producing products in excess of the required quantity [1]. This is because inventory data management is still done manually and sales data is not analyzed so that there are products piling up because they are not selling and there are empty items [2]. One method that can be used is data mining techniques. Data mining is an effective way to obtain information for decision making, one of which is production planning decisions.

The approach to alleviate this issue involves employing the method of clustering. A cluster refers to a group of data objects sharing similarities within the same cluster and distinctiveness from objects found in other clusters. K-Means is an algorithm commonly used in the clustering process [3]. In this research, a rule or guide was developed that contains the relationship of interrelated items using the FP-Growth algorithm. Through this research, it is hoped that researchers can increase sales through technology-based production planning strategies [4].

Numerous prior research endeavors have utilized the K-Means algorithm, primarily for the purpose of categorizing items. For instance, a study applied this algorithm to group Telkomsel card sales regions into three clusters, delineated as high, medium, and low. This grouping served as the foundation for devising promotional procurement strategies [5]. Research was conducted to map technological capabilities by looking at the Davies

Copyright © Acta Logistica, www.actalogistica.eu

Bouldin Index (DBI) value as a basis for measuring cluster performance [6]. Furthermore, studies employing clustering techniques on retail sales data, categorizing them into high, medium, and low clusters, are utilized to formulate stock inventory strategies [7]. Several studies have been carried out, including research conducted by [8] suggests integrating information about customers' physical activity into conventional recommendation systems. Emphasis is placed on providing recommendations tailored to the individual who is jogging or running. These user categories were determined based on the most significant workout-related criteria. Using Orange software, various clustering and classification models were experimented with, leading to the establishment of fundamental criteria for segmentation. The research reveals the feasibility of achieving a fairly precise user segmentation and suggests avenues for refinement. The integration of diverse knowledge models detailing customer behavior and attributes from multiple viewpoints has the potential to yield valuable and distinctive marketing insights [9]. identify and assess the attitudes of the Polish business world towards sustainable entrepreneurship. The findings obtained are juxtaposed with independent assessments, thereby revealing different approaches to business operations. Through cluster analysis of the survey results, it is clear that the Polish SME sector can be categorized into five different groups, each of which shows different approaches and levels of commitment to sustainable development. In addition, the study concludes that Polish companies prioritize social aspects over environmental considerations. Research conducted by [10] Introduces a method to automatically recommend distance measurements for clustering algorithms. The recommendation process involves steps in the form of extracting metadata, which includes collecting meta features and identifying meta targets; building recommendation models using metadata; and suggests distance measures for new datasets through recommendation models. Two different types of meta targets and meta learning techniques are used to address potential variations in user needs. In contrast to prior studies, this research distinguishes itself by employing the K-Means Algorithm to group sales data, generating clusters based on the distinction between high-performing and low-performing sales. Subsequently, the study proceeds with the application of the FP-Growth Algorithm to examine consumer purchasing patterns associated with less popular products. This analysis serves as a foundation for decision-making in the production planning department. This serves as the foundation for conducting this research, aiming to ensure the optimal execution of production planning through the implementation of both the K-Means Algorithm and the FP-Growth Algorithm.

2 Methodology

2.1 Production planning

Production planning is an activity of evaluating past and present facts as well as anticipating future changes and trends to determine appropriate production strategies and schedules in order to achieve targeted goals effectively and efficiently. According to [11] the objectives of production planning and control are:

1. Ensure that the company can produce effectively and efficiently.
2. Ensure that the company can use capital as optimally as possible and can dominate a wide market.
3. Forecast product demand expressed in the number of products as a function of time.
4. Monitor actual demand, compare it with previous demand forecasts and revise these forecasts if deviations occur.
5. Determine the economical order size for the raw materials to be purchased.
6. Monitor inventory levels, compare them with the inventory plan, and revise the production plan at the specified time.
7. Make detailed production schedules, assignments, and machine and labor assignments.

Data mining encompasses a set of procedures designed to extract valuable, previously unknown information from a compiled database, transforming raw data into meaningful insights. The acquired information is derived through the extraction and identification of crucial patterns that contribute to business decision-making. In simpler terms, data mining is commonly described as the process of sifting through vast amounts of data to uncover relevant information. It is also commonly known as Knowledge Discovery in Database (KDD) [12].

2.2 Data mining

Data mining involves a sequence of procedures aimed at manually extracting hidden information value from a gathered database, with the goal of transforming this data into valuable insights. The information gathered is acquired through the extraction and identification of significant patterns essential for informed business decision-making. Put simply, data mining is often described as the process of distilling information from extensive datasets. Another term used to refer to data mining is Knowledge Discovery in Database (KDD) [13]. The initial category of open-source data mining tools, such as WEKA and Rapid Miner, is coded using the JAVA language. These platforms are capable of handling a variety of standard data mining tasks and include the implementation of scalable algorithms for data analysis and extraction [14].

Copyright © Acta Logistica, www.actalogistica.eu

2.3 Clustering

Clustering is a method used to discover and categorize data based on shared characteristics. In data mining, there are two main approaches to grouping data: hierarchical clustering and non-hierarchical clustering. As an unsupervised learning method, clustering partitions a dataset into clusters, with data within each cluster displaying comparable characteristics. Different forms of clustering encompass Hierarchical clustering, Partitioning methods, Density-based clustering, Constraint-based clustering, Fuzzy clustering, and Distribution-based clustering [15]. Hierarchical clustering is a method of classifying data that starts by grouping together two or more objects with close proximity. This process is then repeated, encompassing additional objects that share the next level of proximity. The repetition continues until a cluster structure is established, resembling a tree with a distinct hierarchy (level) among objects, ranging from the most similar to the least similar. Conversely, the non-hierarchical clustering methodology commences by defining the preferred number of clusters (two clusters, three clusters, etc.). Once the number of clusters is established, the clustering process unfolds without adhering to a hierarchical structure. This method is commonly denoted as K-Means Clustering [16].

2.4 K-Means algorithm

K-Means is a non-hierarchical technique for classifying data, aiming to partition the data into two or more groups. This approach segregates data into multiple groups, ensuring that data with similar attributes are assigned to the same group, while data with distinct attributes are allocated to other groups [17]. As mentioned in [18], the procedures for implementing clustering through the K-Means method are outlined as follows:

1. Establish the desired number of cluster centers (*k*) to be created.
2. Randomly initialize (*k*) cluster centers by selecting them from the available data.
3. Compute the distance from every object to each centroid using the Euclidean Distance formula, aiming to identify the closest distance from each centroid data, as expressed in equation (1).

$$D(i,j) = \sqrt{\left(X_{1i} - X_{1j}\right)^2 + \left(X_{2i} - X_{2j}\right)^2 + \cdots + (X_{ki} - X_{kj})^2 \cdots (1)} \quad (1)$$

Note:
$D(i,j)$ - document point,
X_{ki} - criteria data,
X_{kj} - centroid of the cluster *k-j*.

4. Categorize each data point based on its proximity to the centroid, considering the smallest distance.
5. Revise the centroid values, determining the new centroid by calculating the average of the respective cluster, as indicated in equation (2).

$$\mu j(t+1) = \frac{1}{Nsj} \sum j \in sj\ Xj \quad (2)$$

Note:
$\mu j(t+1)$ - new centroid in the th iteration ($t + 1$),
Nsj - the amount of data in cluster *sj*,

6. Iterate through steps 2 to 5 until each cluster attains a stable value.

2.5 Association rules

Association is a technique utilized in the data mining process to extract association rules, revealing connections between item combinations or relationships among attributes. Association rules are procedures for finding relationships between items in a predefined data set. Data mining technology based on association rules is a crucial method for discovering knowledge. Its objective is to identify correlations among characteristic variables and determine their level of proximity within extensive datasets. This technology aims to unveil potential, concealed, and valuable information in samples, thereby offering significant decision support. The findings of prior research suggest that data mining based on association rules can uncover insights that are difficult to extract using conventional statistical approaches. This approach proves to be practical and effective [19].

Association is a method for examining connections between specified pairs of items. It involves deriving associative rules from two metrics: support, which indicates the percentage of item combinations in the data, and confidence, which signifies the accuracy strength of the relationship between items. The fundamental methodology of association analysis comprises two stages designed to identify all association rules that satisfy the minimum support and minimum confidence criteria [20].

In the exploration phase for the most frequent pattern, the FP-Growth Algorithm is employed to discover combinations of items that fulfill the minimum support value criteria in the database. In the context of the support formula in data mining, P refers to an itemset or set of items (a collection of items or attributes) in a transaction or data set. the support value of an item obtains the calculation of the support value of two items written in equation (3).

$$Support\ (P) = \frac{\text{Number of Transactions Containing P}}{\text{Total Transactions}} \times 100\% \quad (3)$$

Support (P) is the level of support for itemset P. Number of transactions containing Number of transactions containing P is the number of transactions in which itemset A exists.

Transaction totals Transaction totals are the total number of transactions in the data set. In the context of the

Copyright © Acta Logistica, www.actalogistica.eu

Acta logistica - International Scientific Journal about Logistics
Volume: 11 2024 Issue: 2 Pages: 233-243 ISSN 1339-5629

support formula for association rules, P and B refer to the two itemsets or sets of items that are being evaluated in a transaction or dataset (4).

$$Support\ (\mathrm{P} \cap \mathrm{B}) = \frac{\text{Number of Transactions Contains P and B}}{\text{Total Transactions}} \times 100\% \tag{4}$$

Support (P∩B) is the level of support for the combination of itemsets P and B.

The number of transactions containing P and B is the number of transactions containing P and B is the number of transactions in which both itemsets P and B occur together. Total transactions is the total number of transactions in the data set.

After identifying all high-frequency patterns, the next step is to search for association rules that satisfy the minimum confidence criteria, indicating how frequently an item appears concurrently with a combination of other items. Associative rule A → B The probability that B appears when P also appears can be obtained from equation (5).

$$Confidence\ \mathrm{Q}\,(\mathrm{P} \mid \mathrm{A}) = \frac{\text{Number of Transactions Contains P and B}}{\text{Total Transactions Contain P}} \times 100\% \tag{5}$$

2.6 FP-Growth Algorithm

The FP-Growth algorithm, belonging to the associative data mining approach, is instrumental in efficiently handling frequent itemsets through the utilization of the FP-Tree structure. This algorithm is founded on the FP-tree concept. FP-Tree construction involves linking data from individual transactions along specific paths. This linkage allows transactions sharing common items to overlap on paths. The efficiency of the data structure layout process increases with a greater number of transactions featuring the same items. After the FP-Tree is formed, then carry out the three main stages of the FP-Growth Algorithm, namely [21]:

1. Stage of creating a conditional pattern base.
2. Stage of generating a conditional FP-Tree.
3. Stage of generating frequent items.

3 Result and discussion

3.1 Sales transaction data

In this study, the dataset utilized comprises sales transaction records spanning from January to December 2022. The objective of this research is to categorize products into two criteria, namely slow-moving and fast-moving items, and derive patterns from the formed clusters.

3.2 Data processing

In this study, the sales transaction data for XY Bakery covering a one-year period, from January to December 2022, is employed. Two methodologies, the K-Means Clustering Algorithm and the FP-Growth Algorithm, are applied to process the data. The purpose of employing these methods is to obtain production planning outcomes in alignment with the data mining approach. The data, organized through the utilization of the K-Means Clustering Algorithm, serves as foundational data to analyze consumer purchasing patterns through association rules. Both methodologies involve multiple stages of data mining, starting with the data selection process that focuses on gathering sales data specifically for products containing bread items. Based on the data selection results, the data cleaning stage continues, namely eliminating noise by removing duplicate data and correcting data errors such as transactions that only contain 1 type of item.

3.2.1 Bread grouping based on the K-Means Clustering Algorithm

The initial step in the K-Means Clustering Algorithm involves determining the number of datasets and conducting an initial performance test to identify the optimal k value. The selection of the most suitable number of classes is assessed by observing the smallest value on the Davies-Bouldin index. The Davies-Bouldin Index serves as a technique for assessing the appropriateness of the cluster count [22]. Based on the performance test results, 3 classes were formed. However, after clustering there are classes that only have one member. This category is deemed to lack a substantial impact on the configuration of clusters. Consequently, only two categories (k=2) were employed in this study. Out of this dataset, 59 samples were extracted that satisfied the research criteria for implementing the K-Means algorithm. The steps in the K-Means Clustering algorithm are outlined below:

1. Establish the initial centroid value. There are two clusters: C1 represents slow-moving product data, and C2 represents fast-moving products.
2. Determine the initial centroid by taking data randomly. The centroid is determined based on bread sales transaction data for 12 months, which is classified based on the name of the bread item and the bread sales period. Table 1 shows sales transaction data that has been randomly selected for centroids.

Copyright © Acta Logistica, www.actalogistica.eu

Table 1 Ccentroid selection

No	Item	Jan (pcs)	Feb (pcs)	Mar (pcs)	Apri (pcs)	May (pcs)	Jun (pcs)	Jul (pcs)	Ags (pcs)	Sept (pcs)	Oct (pcs)	Nov (pcs)	Dec (pcs)
1	Abon	276	134	193	106	143	166	177	223	245	255	296	256
2	Abon Gulung	478	337	400	211	288	281	292	346	724	256	486	361
*	****	*	*	*	*	*	*	*	*	*	*	*	*
11	Choco Ball W	6	4	9	1	1	4	1	1	1	20	8	10
*	****	*	*	*	*	*	*	*	*	*	*	*	*
59	Wool Roll Bread Tuna	57	41	48	83	74	74	60	59	70	55	98	62

Centroid 1 (11^{th} data): {6, 4, 9, 1, 1, 4, 1, 1, 1, 20, 8, 10}
Centroid 2 (2^{nd} data): {478, 337, 400, 211, 288, 281, 292,346, 724, 256, 486, 361}

After choosing the initial centroid value, the Euclidean distance theory is applied to calculate the proximity of data to the centroid. Each data will be calculated as the closest distance to each centroid that was determined in the previous stage using equation 1. In the first iteration, category C1 produces 51 members. Category C2 produces 8 members. Then proceed with calculating the 2nd iteration.

1. Update the centroid value by calculating the average of the cluster in question using equation 2.
2. The calculation procedure remains consistent with the previous step. If the cluster positions in the new iteration match those of the previous iteration, the process stops. However, if they differ, the process proceeds to the next iteration.

In the second iteration, C1 yields 45 members, while C2 produces 14 members. As the cluster positions in the second iteration differ from those in the previous iteration, the process advances to the third iteration, requiring an update to the centroid value.

In the third iteration, C1 generated 44 members, while C2 produced 15 members. As the cluster positions in the third iteration differ from those in the previous iteration, the process proceeds to the fourth iteration, necessitating an update to the centroid value.

In the fourth iteration, C1 generated 44 members, and C2 produced 15 members. The cluster positions in the fourth iteration are identical to those in the previous iteration, so the process is halted. A comparison of the index for the number of data objects in each cluster is illustrated in Table 2.

Table 2 Comparison of the number of objects for each cluster

Iteration process	Number of objects C1	Number of objects C2
Iteration 1	51	8
Iteration 2	45	14
Iteration 3	44	15
Iteration 4	44	15

The calculation ends in the 4th iteration because it has the same results as the results of the 3rd iteration, causing the ratio in the 4th iteration to remain unchanged. The data group is then declared convergent or considered optimal. The next stage is to carry out testing using Rapidminer to see consumer purchasing patterns over the year. The figure 1 is the sales graph display for XY bread products. Figure 1 consumer purchasing patterns within a year, the x-axis shows the demand period and the y-axis shows the amount of consumer purchases. The blue line graph shows slow moving bread products and the red line shows fast moving bread products. Group 1 as slow moving products has 44 types of items and group 2 as fast moving products has 15 types of items. Based on this graph, it can be concluded that XY bakery is experiencing losses due to the large number of overstock products.

Copyright © Acta Logistica, www.actalogistica.eu

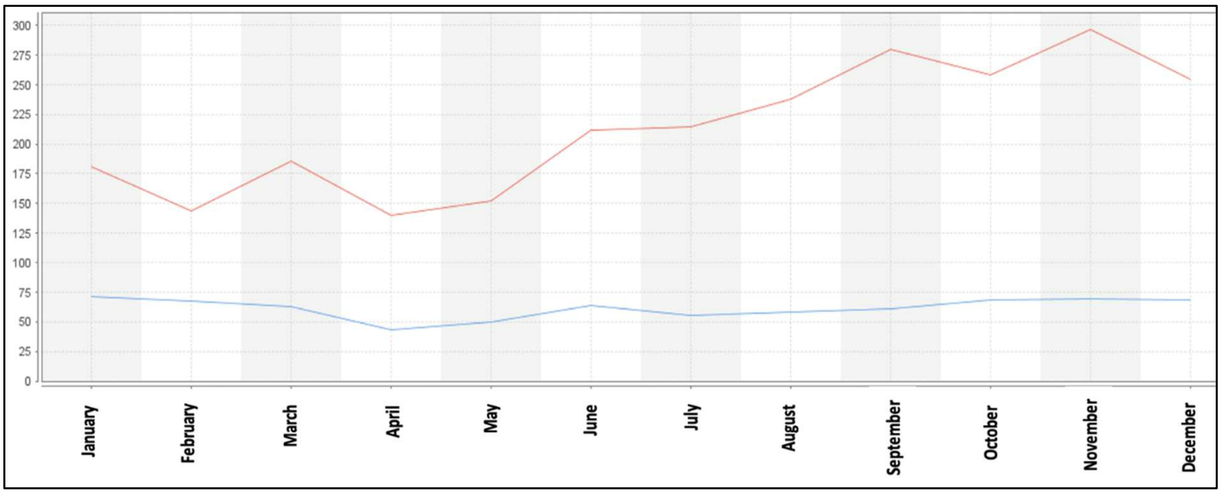

Figure 1 Graphic display based on testing using RapidMiner

3.2.2 Determining consumer purchasing patterns using the FP-Growth Algorithm

The identification of consumer purchasing patterns is derived from the categorization of slow-moving and fast-moving data. When establishing consumer purchasing patterns, the data mining stages parallel those of the K-Means Clustering method. Consequently, the outcomes from K-Means serve as the foundational material for processing FP-Growth data. The subsequent steps outline the stages of the FP-Growth Algorithm in handling sales transaction data [23].

3.3 Frequency of each item and support value selection

This research sets a minimum support value of 2%. After that, a data selection process was carried out and deleted data that had a support value below 2% with a total transaction of 58,802. The process of forming 1 itemset with a minimum support value of 2% uses equation 3 to obtain a support value as in table 2 which meets the minimum support of 2%:

Table 2 Pattern 1 itemset that meets minimum support

Item type	Frequency	*Support*	Percentage of support value
Tuna Fish Buns	901	(901/58802) x 100%	2%
Wool Roll Bread Coklat	1015	(1015/58802) x 100%	2%
Abon	2015	(2015/58802) x 100%	3%
***	**	***	**
Blueberry	2086	(2086/58802) x 100%	4%
Vanila Almond	901	(901/58802) x 100%	4%
Abon Gulung	3605	(3605/58802) x 100%	6%

After scanning items that have a support count frequency = 2%, the number of items will be entered into the FP-Tree. In this procedure, items failing to meet the minimum support requirement will be eliminated due to their negligible impact, and a data cleaning process will be conducted on transaction data containing only one item. The subsequent computation is executed to assess the occurrence frequency of each item after the data cleaning process. The table below displays the frequency of each frequent item in individual transactions, arranged in descending order based on their highest frequencies.

Table 1 Frequency of item appearance after data cleaning process

Item Name	Frequency of appearance
Abon Gulung	404
Abon	254
Blueberry	251
Brown Sugar	250
**	**
Pisang Keju	118
Cheese Roll	115
Tuna Fish Buns	113
Sosis Sate	109

The table above shows the frequency of appearance of items after going through the data cleaning process.

Copyright © Acta Logistica, www.actalogistica.eu

Acta logistica - International Scientific Journal about Logistics
Volume: 11 2024 Issue: 2 Pages: 233-243 ISSN 1339-5629

3.4 *FP-Tree formation*

FP-Tree is formed based on item categories which have been sorted based on priority for each transaction. The formation of this FP-Tree is based on the researcher's policy of compiling transaction flows based on transactions containing slow moving products in accordance with the research objective, namely to optimize production by minimizing wasted product.

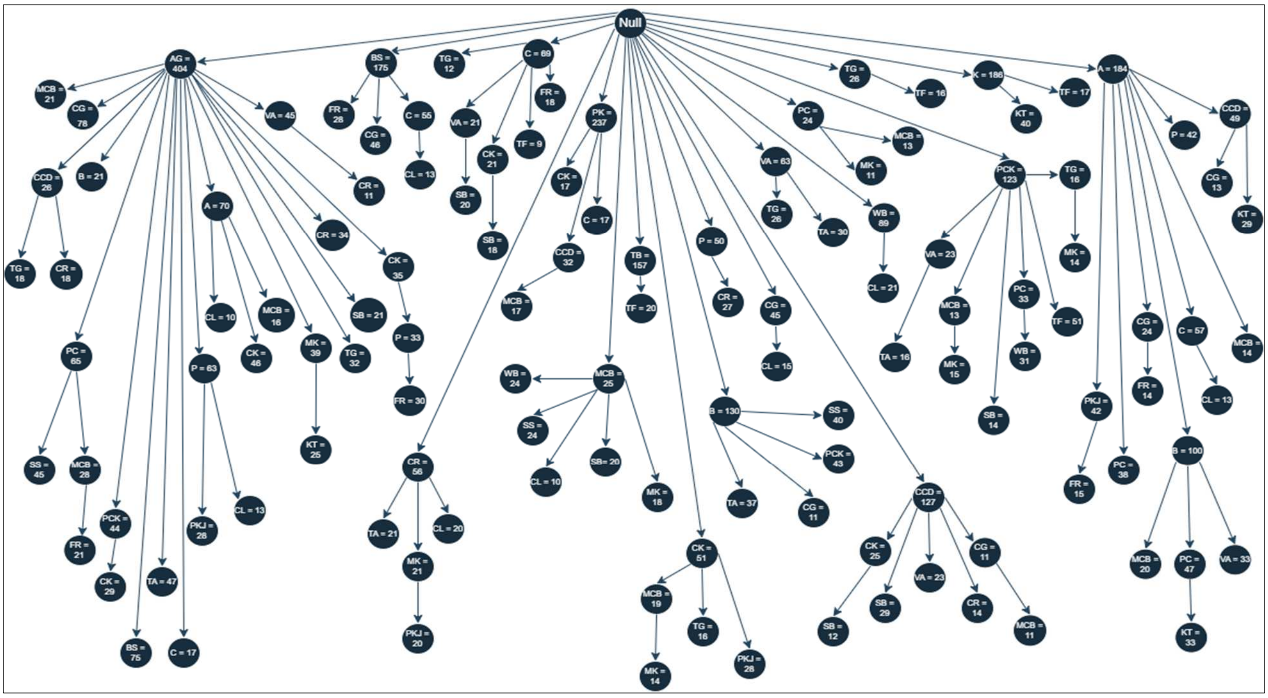

Figure 1 FP-Tree on reading all transactions

Generation of FP-Tree from all transaction data containing slow moving category items. Every node in the FP-Tree includes the initials of the item's name coupled with a support counter. Following this, a quest will be undertaken using the FP-Growth algorithm process to unearth noteworthy frequent itemsets. FP-Growth comprises three phases:

1. Conditional Pattern Base
   The Conditional Pattern Base is a subdatabase incorporating the prefix path and suffix pattern derived from the FP-Tree constructed in the preceding phase, beginning with the item featuring the lowest support count. Here are the outcomes of generating the conditional pattern base.

Table 2 Generation of Conditional Pattern Base

Suffix	Generation of Conditional Pattern Base
SS	(AG, PC : 45), (MCB :24), (B : 40)
TF	(C :9), (TB : 20), (K : 17), (PCK : 51), (TG : 16)
CL	(AG, P : 13), (BS, C : 13), (CR : 20), (MCB : 10), (CG : 15), (WB : 21), (A,C : 13)
PKJ	(CR, MK : 20),(CK : 280), (A : 42)
**	*****
BS	(AG : 75)
B	(AG : 21), (A : 100)
A	(AG : 70)
AG	-

2. Conditional FP-Tree
   During this step, the support counts for each item in each conditional pattern base are summed. For each category with a support count greater than or equal to the predefined minimum support count (25), it will be revived with a conditional FP-Tree. Categories with a support count less than the minimum support count will not be resurrected. Below is a table illustrating the results of the conditional FP-Tree search.

Copyright © Acta Logistica, www.actalogistica.eu

Table 3 Generation of Conditional FP-Tree

Suffix	**Generation of Conditional Pattern Base**
SS	(AG : 45), (B : 40), (PC :45), (AG, PC : 45)
TF	(PCK : 51)
PKJ	(A : 42), (CK : 28)
FR	(AG : 30), (A : 29), (BS : 28), (CK : 30), (P : 30), (AG, CK, P : 30)
**	*****
BS	(AG : 75)
B	(A : 100)
A	(AG : 70)
AG	-

3. Frequent Itemset
In this step, the arrangement of items established by each conditional FP-Tree is implemented. If it doesn't shape a single path, the FP-Growth generation is conducted recursively. The results of the search for frequent item sets are outlined below:

Table 4 Frequent Itemset Formation

Suffix	**Generation of Conditional Pattern Base**
SS	(AG, SS : 45), (B, SS : 40), (PC, SS :45), (AG, PC, SS : 45)
TF	(PCK, TF : 51)
PKJ	(A, PKJ : 42), (CK, PKJ : 28)
FR	(AG, FR : 30), (A, FR : 29), (BS, FR : 28), (CK, FR : 30), (P, FR : 30), (AG, CK, P, FR : 30)
**	*****
BS	(AG, BS : 75)
B	(A, B : 100)
A	(AG, A : 70)
AG	-

Table 7 is the result of determining consumer purchasing patterns which provides information as a reference in determining marketing strategies. Following the creation phase, the subsequent step involves establishing the minimum support and minimum confidence.

3.5 *Determining support values and confidence values using Association Rules*

The outcomes of this association procedure furnish insights into items acquired together, determined by calculating the support and confidence values for each itemset using equations 4 and 5. Adhering to the stipulations of a minimum 3% support and 30% confidence, 13 rules are identified, classifying them as robust association rules. It can be seen in the following Table 8.

Table 5 Association Rule with Support Value and Confidence Value

Frequent	***Support***	***Confidence***
PCK ; TF	3%	45%
A; KT	3%	49%
AG;TG	3%	34%
AG;CR	3%	39%
AG;P	5%	51%
AG;MCB	3%	33%
AG;PC	3%	31%
A;PC	4%	41%
AG;VA	11%	100%
AG;CK	6%	49%
AG;CG	4%	34%
AG;BS	4%	30%
A;B	5%	40%

Copyright © Acta Logistica, www.actalogistica.eu

Volume: 11 2024 Issue: 2 Pages: 233-243 ISSN 1339-5629

According to the Support Values and Confidence Values of the Association Rules, there are 13 rules that satisfy the association rule criteria, representing the conclusive rules for analyzing the XY bakery sales transaction dataset. The 13 association rules can be explained as follows :

1. If you buy the “Pisang Coklat Keju” (PCK) variant of bread, then there is a possibility of buying the “Tuna Fish Buns” (TF) variant with support of 3% and confidence of 45%
2. If you buy the “Abon” bread variant (A), there is a possibility of buying the “Keju Tabur” variant (KT) with support of 3% and confidence of 49%.
3. If you buy the “Abon Gulung” bread (AG) then you are likely to buy the “Tawar Gandum” (TG) variety with support of 3% and confidence of 34%.
4. If you buy the “Abon Gulung” (AG) variant of bread, there is a possibility of buying the “Cinnamor Roll” (CR) variant with support of 3% and confidence of 39%
5. If you buy the “Abon Gulung” bread variant (AG) then the possibility of buying the “Pizza” variant bread (P) is with support of 5% and confidence of 51%.
6. If you buy the “Abon Gulung” bread (AG) variant, there is a possibility of buying the “Mexican Coffee Buns” (MCB) variant with support of 3% and confidence of 33%.
7. If you buy the “Abon Gulung” (AG) bread variant, there is a possibility of buying the “Pisang Coklat” (PC) variant with support of 3% and confidence of 31%.
8. If you buy the “Abon” bread variant (A), there is a possibility of buying the “Pisang Coklat” variant (PC) with support of 5% and confidence of 41%.
9. If you buy the “Abon Gulung” bread (AG) then you are likely to buy the “Vanilla Almond” (VA) bread with 11% support and 100% confidence.
10. If you buy the “Abon Gulung” bread (AG) then you are likely to buy the “Coklat Keju” (CK) variant with support of 6% and confidence of 49%.
11. If you buy the “Abon Gulung” bread (AG) then the possibility of buying the “Coklat Gulung” (CG) variant is with support of 4% and confidence of 34%.
12. If you buy the “Abon Gulung” (AG) bread variant, there is a possibility of buying the “Brown Sugar” (BS) variant with support of 4% and confidence of 30%.
13. If you buy the “Abon” bread variant (A), there is a possibility of buying the “Blueberry” variant of bread (B) with support of 5% and confidence of 40%.

After consumer purchasing patterns are formed, the next stage is to carry out testing using RapidMiner. From the test results using RapidMiner, there are several patterns that are different from the results of calculations carried out manually. This is because in manual calculations the researcher prioritizes transactions with itemsets containing slow moving products, whereas in the RapidMiner test these are not input in processing the data. Figure 3 shows the test results using RapidMiner.

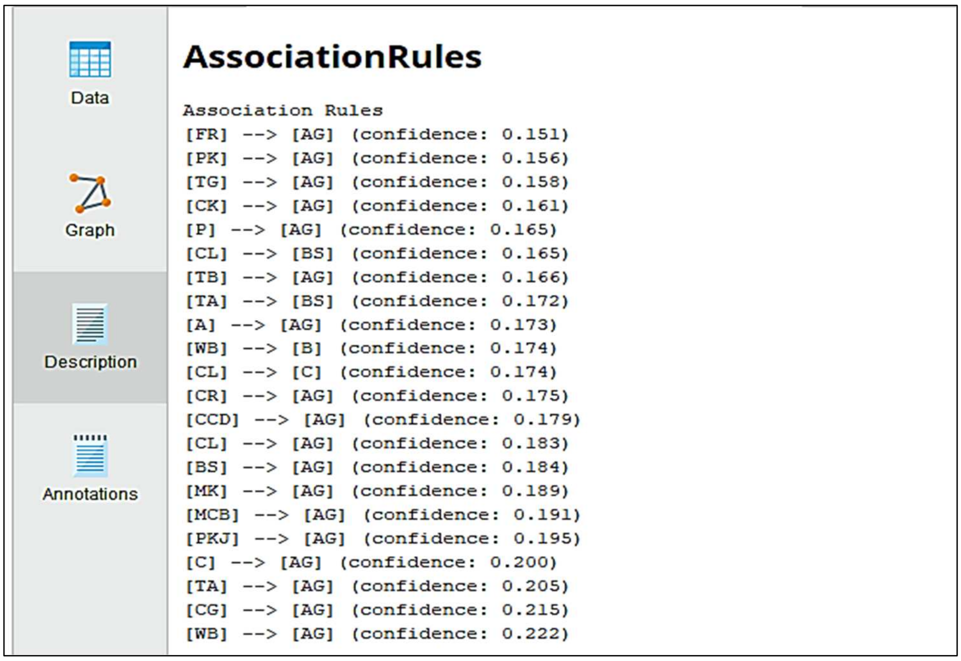

Figure 2 Results of Consumer Purchasing Patterns Based on RapidMiner Testing

Copyright © Acta Logistica, www.actalogistica.eu

3.5.1 Production planning based on the results of the K-Means Algorithm and FP-Growth Algorithm

According to the data mining analysis, it is evident that the sales data for bread at the XY bakery indicates a higher proportion of slow-moving products compared to fast-moving products. The data outcomes reveal that XY bakery faced significant losses attributed to a considerable overstock of products, indicating that the bakery did not achieve its sales target. In response to these challenges, the next step involves formulating a marketing strategy using the FP-Growth method.

Following the identification of patterns derived from the K-Means algorithm's grouping outcomes, the subsequent step involves implementing association rules to elucidate the connections among items in a dataset and assess the likelihood of simultaneous bread purchases. This known consumer purchasing pattern can be the basis for decision making in the field of production planning. In this case the researcher does not determine the actual value but can provide a pattern of 13 rules which produces a strong level of association. Through information from this data, especially for fast moving products, shop owners can improve inventory planning by analyzing demand data and market trends. For slow moving products, shop owners can arrange the placement of goods or create product bundling with best-selling products. Store owners can also design promotions for these item combinations by providing discounts or offering giveaways.

4 Conclusions

Through the application of the K-Means Clustering Algorithm, the formation of two distinct classes, namely slow-moving and fast-moving, was realized. The computational process extended up to the 4th iteration, where group 1, representing slow-moving products, encompassed 44 distinct types of bread items, while group 2, denoting fast-moving products, comprised 15 different items. These calculations were conducted both manually and using the Rapidminer software.

Based on the purchasing pattern that has been formed, the support value and confidence value are then determined to obtain a pattern that has a close relationship. Patterns that meet the minimum support and minimum confidence values as the final dataset rules include :

1. If you buy the “Pisang Coklat Keju” (PCK) variant of bread, then there is a possibility of buying the “Tuna Fish” buns (TF) variant with support of 3% and confidence of 45%.
2. If you buy the “Abon” bread variant (A), there is a possibility of buying the “Keju Tabur” variant (KT) with support of 3% and confidence of 49%.
3. If you buy the “Abon Gulung” bread (AG) then the possibility of buying the “Tawar Gandum” (TG) variant with support is 3% and confidence is 34%.

The employed method aids in offering insights and details concerning product availability or stock fluctuations of frequently purchased items by consumers. This information serves as a guide for anticipating product supply in accordance with demand, mitigating the risk of supply shortages.

The application of the K-Means algorithm can help develop strategies in determining stock of goods. It is hoped that this method can be combined again with other algorithms to get better results and the data that has been obtained shows that the bakery has not reached BEP (Break Even Point) in its sales so that a stage is needed continued by carrying out marketing strategies using the FP-Growth method for subsequent research opportunities.

Acknowledgement

The author would like to thank Hasanuddin University Engineering for providing financial support through the LBE (Laboratory Base Education) scheme with research contract number 11367/UN7.2/PM.01.01/2023. In addition, the author would like to thank the Manufacturing Systems Laboratory, Department of Industrial Engineering, Hasanuddin University, which has actively participated during this research process.

References

[1] JURÍK, L., HORŇÁKOVÁ N., DOMČEKOVÁ, V.: The Application of SMED Method in The Industrial Enterprise, *Acta logistica*, Vol. 7, No. 4, pp. 269-281, 2020. https://doi.org/10.22306/al.v7i4.189

[2] IMROM, M., HASANAH, U., HUMAIDI, B.: Analysis of Data Mining Using K-Means Clustering Algorithm for Product Grouping, *International Journal of Informatics and Information Systems*, Vol. 3, No. 1, pp. 12-22, 2020. https://doi.org/10.47738/ijiis.v3i1.3

[3] MIFTAKHUL, M.S., PRIHANDOKO.: Penerapan Algoritma K-Means dan Cure dalam Menganalisa Pola Perubahan Belanja dari Retail ke E-Commerce, *Jurnal Energy*, Vol. 7, No. 2, pp. 44-49, 2017. (Original in Indonesian)

[4] CAESAR, F.X B., SOMYA, R.: *Analisis Minat Beli Produk pada Toko Oleh-oleh Khas Surabaya dengan Algoritme FP-Growth*, Seminar Nasional Dinamika Informatika Universitas PGRI, Yogyakarta, pp. 5-10, 2021. (Original in Indonesian)

[5] HANDOKO, S., FAUZIAH., E.T.E., HANDAYANI: Implementasi Data Mining untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering, *Jurnal Ilmiah Teknologi dan Rekayasa*, Vol. 25, No. 1, pp. 76-88, 2020. (Original in Indonesian)

[6] BRAVO, M.C.M., CHALEZQUER, C.C., PUCHE, J.S.: Meta-Framework of Digital Literacy: A Comparative Analysis of 21st-Century Skills Framework, RLCS, *Revista Latina de Communicacion Social*, Vol. 79, pp. 76-110, 2021. https://doi.org/10.4185/RLCS-2021-1508

Copyright © Acta Logistica, www.actalogistica.eu

[7] NEGARA, I.S.M., PURWONO, ASHARI, I.A.: Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means, *Journal of Infromation Technology and Computer Science*, Vol. 6, No. 3, pp. 153-160, 2021. (Original in Indonesian)
[8] PAWELOSZEK, I.: Customer Segmentation Based on Activity Monitoring Applications for The Recommendation System, *Procedia Computer Science*, Vol. 192, No. 3, pp. 4751-4761, 2021. https://doi.org/10.1016/j.procs.2021.09.253
[9] BADJOR, P., PAWELOSZEK, I.,FIDLEROVA, H.: Analysis and Assessment of Sustainable Entrepreneurship Practices in Polish Small and Medium Enterprises, *Sustainability*, Vol. 13, No. 7, pp. 1-28, 2021. https://doi.org/10.3390/su13073595
[10] AL-CHALABI, H.H., JASIM, M.N.: *Food Recommendation System Based on Data Clustering Techniques and User Nutrition Records*, International Conference on New Trends in Information and Communications Technology Applications, Vol. 1764, 16-17 November, Baghdad, 2022. https://doi.org/10.1007/978-3-031-35442-7_8
[11] EUNIKE, A., SETYANTO, N.W., YUNIARTI, R., HANDALA, I., LUKODONO, R.P., FANANI, A.A.: *Perencanaan Produksi dan Pengendalian Persediaan*, Malang, UB Press, 2018. (Original in Indonesian)
[12] ATEAGA, V.F., ZARATE, C.A.T., FRESAN, A., CASTRO, T.B.G., ROJOP, I..E.J., NARVAEZ, L.L., DIAZ, Y.H.: Association Between Completed Suicide and Environmental Temperature in a Mexican Pupulation, Using The Knowledge Discovery in Database Approach, *Computer Methods and Programs in Biomedicine*, Vol. 135, pp. 219-224, 2016. https://doi.org/10.1016/j.cmpb.2016.08.002
[13] VUCETIC, M., HUDEC, M., BOZILOVIC, B.: Fuzzy Functional Dependencies and Linguistic Interpretations Employed in Knowledge Discovery Tasks from Relational Database, *Engineering Application of Artificial Intelligence*, Vol. 88, pp. 1-15, 2020. https://doi.org/10.1016/j.engappai.2019.103395
[14] BIAO, Z.Z.: Design and Realization of Data Mining Simulation and Methodological Models, *Journal of King Saudi University-Science*, Vol. 35, No. 10, pp. 1-9, 2023. https://doi.org/10.1016/j.jksus.2023.102964
[15] SHI, H., PENG, Q., XIE, Z., WANG, J.: A Semi-Supervised Hierarchical Ensemble Clustering Framework Based on A Novel Similarity Metric and Stratified Feature Sampling, *Journal of King Saud University - Computer and Information Science*, Vol. 35, pp. 1-11, 2023. https://doi.org/10.1016/j.jksuci.2023.101687
[16] SANTOSO, S.: *Statistik Multivariat Konsep dan Aplikasi dengan SPSS*, Jakarta, PT Elex Media Komputindo, 2010. (Original in Indonesian)
[17] HASYRIF, S.Y., RISMAYANI, SYAM, A.: *Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare di Kota Makassar*, Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi, Makassar, Vol.8, No. 1, pp. 73-82, 2019. (Original in Indonesian)
[18] SUNITA, D.M., PRADIP, M.J.: Classification Technique and Its Combination with Clustering and Association Rule Mining in Educational Data Mining - A Survey, *Engineering Applications of Artificial Intelligence*, Vol. 122, pp. 1-56, 2023. https://doi.org/10.1016/j.engappai.2023.106071
[19] ZHU, L., LIU, J.: The Decision Supports for Male Migrant Workers' Physical Features at Different Stages of Physical Exercise Behavior by Association Rules Based Data Mining Technology, *Procedia Computer Science*, Vol. 166, pp. 448-455, 2020. https://doi.org/10.1016/j.procs.2020.02.066
[20] ATTURROHMA, I.: *Penentuan Tata Letak Barang Dagangan Berdasarkan Data Transaksi Penjualan Harian Menggunakan Algoritma Apriori*, Seminar Nasional Rekayasa Teknologi Informasi, Serang, Vol. 1 pp. 155-168, 2018. (Original in Indonesian)
[21] WIBOWO, A.R., JANANTO, A.: Implementasi Data Mining Metode Asosiasi Algoritma FP-Growth Pada Perusahaan Ritel, *Teknologi Informasi dan Komunikasi*, Vol. 10, No. 2, pp. 200-212, 2020. https://doi.org/10.35585/inspir.v10i2.2585 (Original in Indonesian)
[22] MUGHNYANTI, M., EFENDI, S., ZARLIS, M.: *Analysis of Determining Centroid Clustering X-means Algorithm with Davies-bouldin Index Ecaluation*, IOP Conference Series: Materials Science and Engineering, Medan, Vol. 725, pp. 1-6, 2020. https://doi.org/10.1088/1757-899X/725/1/012128
[23] NURAROFAH, E., HERDIANA, R., NURIS, D.N.: Penarapan Asosiasi Menggunakan Algoritma FP-Growth Papda Pola Transaksi Penjualan Di Toko Roti, *JATI Jurnal Mahasiswa Teknik Informatika*, Vol. 7, No. 1, pp. 353-359, 2023. https://doi.org/10.36040/jati.v7i1.6299 (Original in Indonesian)

Review process
Single-blind peer review process.

Copyright © Acta Logistica, www.actalogistica.eu