Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

Apriori Algorithm to Predict Availability of Beauty Products

Maria Hikmah Hasibuan[1)*], M. Fakhriza[2)]
[1)*2)]Computer Science Department, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
[1)*] mariahikma05@gmail.com, [2)]fakhriza@uinsu.ac.id

ABSTRACT

This study introduces the Apriori algorithm in beauty product availability prediction system as a solution to enhance stock prediction accuracy and mitigate inventory risks in the beauty industry. By applying data mining technology, specifically the Apriori algorithm, Kazana Kosmetik aims to gain insights into consumer purchasing patterns to optimize operations. The research analyzes transaction data to identify key buying patterns and improve stock management strategies. The results reveal seven main purchasing patterns with an average confidence value of 0.414, offering valuable guidance for Kazana Kosmetik in inventory control and marketing tactics. By leveraging data mining techniques, companies like Kazana Kosmetik can streamline sales strategies and enhance customer satisfaction. This research underscores the effectiveness of the Apriori algorithm in predicting beauty product availability and its potential to revolutionize operational efficiency in the cosmetics market.

Keywords: Data Mining; Apriori Algorithm; Beauty Products; Inventory Management; Consumer Behavior

1. INTRODUCTION

The Apriori algorithm is useful for businesses in making decisions on the availability of goods, such as in the business field. It helps determine what cosmetics are bought by consumers and analyze consumer buying habits. As technology continues to develop, more competitors in the business world require developers to increase sales and marketing of products sold. Sales transactions accumulate daily due to increased production, making it difficult for resellers of beauty product sales to control their transaction data and use it as useful information.

Kazana Kosmetik, a company selling various brands of beauty cosmetics, faces challenges in controlling sales transactions and identifying frequently sold items. With 120-200 sales transaction production per day and 840-1000 sales products per month, the store's in-out system results in neglected or out-of-stock products. This method of ordering results in old or sold products being neglected. Additionally, the display of organized products also poses a challenge in increasing product sales.

To address these problems, researchers propose the "Apriori Algorithm in Beauty Product Availability Prediction System." The goal is to predict the availability of beauty products using the Apriori algorithm to improve accuracy in predicting stock and reduce the risk of shortages or excess inventory. By utilizing data mining technology, Kazana Kosmetik can better understand consumer buying habits and improve its operations.

The Apriori algorithm can be an effective solution for beauty product retailers to predict product availability and improve efficiency in the market. By leveraging data mining technology, companies like Kazana Kosmetik can better manage their sales and marketing strategies, ultimately benefiting both the industry and their customers.

2. LITERATURE REVIEW

2.1 Data Mining

Data mining is the process of finding interesting patterns or information in selected data using certain techniques or methods (Setiawan et al., 2023). The techniques, methods, or algorithms in data mining vary widely. The selection of the right method or algorithm is highly dependent on the objectives and the overall Knowledge Discovery in Database (KDD) process.

According to the Gartner Group, data mining is the process of discovering new meaningful relationships, patterns and habits by sorting through large portions of data stored in storage media using pattern recognition technologies such as statistical and mathematical techniques. Data mining is a combination of several disciplines that bring together techniques from machine learning, pattern recognition, statistics, databases, and visualization to handle information retrieval problems from large databases (Urva et al., 2023).

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

Data mining according to David Hand, Heikki Mannila, and Padhraic Smyth from MIT is the analysis of data (usually large data) to find clear relationships and infer them that have not been known before in a way that is currently understood and useful for the owner of the data (Setiawan et al., 2023).

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from various large databases. Data mining is a series of processes to extract added value from a data set in the form of knowledge that has not been known manually.

2.1.1 Knowledge Discovery In Databases (KDD)

(Butar, 2022), defines that the terms data mining and knowledge discovery in database (KDD) to describe the process of extracting hidden information in a large database, data mining and KDD are often used interchangeably. The two terms have different concepts, but are related to each other. One of the stages in the whole KDD process is data mining. As addressed in the following process:

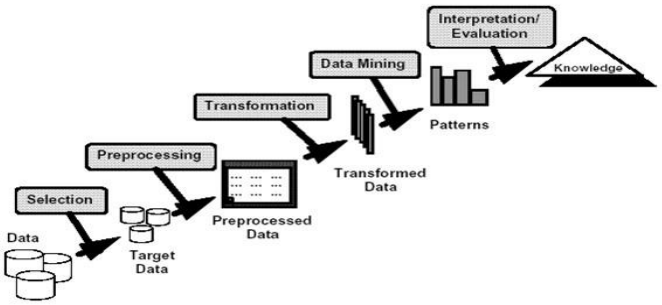

Fig. 1 Data Mining Process Stages

1. Data selection
   Before the information mining stage in KDD, it is necessary to select data from a set of data. What is used for the data mining process is the selected data, stored in a separate archive from the operational database.
2. Pre-processing/cleaning
   Cleaning the data, which is the focus of KDD, is the beginning of the process before performing the data mining process. Removing duplicate data, checking for inconsistent data, and correcting errors in the data are the stages of the cleaning process. The process of "enriching" existing data with other significant data (information) needed for KDD, such as external data or information is a process carried out in enrichment.
3. Transformation Coding
   Transformation coding is a coding technique used in signal processing and data compression to transform the signal domain into a transform domain that is more suitable for compression or analysis. The main goal of transformation coding is to reduce data redundancy and improve representation efficiency.
4. Data Mining
   Data mining uses certain techniques or methods in extracting interesting patterns or information in selected data. The methods involved in data mining are very diverse. The overall goal and process of KDD will determine the appropriate method or algorithm in the overall process.
5. Interpretation/Evaluation
   Displaying the process of a data mining information pattern in a form that is easy to understand. This stage is called interpretation which is part of the KDD process, with an examination of the pattern or information found whether it is different from the facts or hypotheses that existed before.

2.1.2 Data Mining Techniques

Data Mining is divided into several techniques based on the tasks that can be performed, namely: (Urva et al., 2023)

1. Description
   Sometimes researchers and analysts simply want to try to find ways to describe the patterns and trends contained in the data. Descriptions of patterns and trends often provide possible explanations for a pattern or trend.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

2. Estimation
   Estimation is similar to classification, except that the target variable for estimation is numerical rather than categorical. The model is built using complete records that provide the value of the target variable as a predicted value. Then, at the next review, the estimated value of the target variable is made based on the value of the prediction variable.
3. Prediction
   Prediction is almost the same as classification and estimation, except that in prediction the value of the result will be in the future. Some of the methods and techniques used in classification and estimation can also be used (under the right circumstances) for prediction.
4. Classification
   In classification, there are target variable categories. For example, income classification can be separated into three categories, namely high income, medium income, and low income.
5. Clustering
   Using this technique, data is grouped into sets that are comparable to each other based on the similarity of certain attributes. A cluster is a set of records that are similar to each other and have no similarity with records in other clusters.
6. Association.
   The task of association in data mining is to find attributes that appear at one time. This technique is used to find relationships or associations between variables in a data set.

2.2 Apriori Algorithm

The Apriori algorithm is an algorithm used in data mining to find association patterns in data sets (Putra et al., 2023). Association patterns refer to relationships between items in a data set that often appear together. For example, in a cosmetics store, association patterns can help in identifying the set of items that are often purchased together by customers.

Apriori Algorithm uses the knowledge of previously known attribute frequencies to process further information. The Apriori Algorithm determines the candidates that may appear by paying attention to the minimum support and minimum confidence. Support is the value of support or the percentage of combinations of an item in the database. Confidence is the certainty value, namely the strength of the relationship between items in the associative rules formed. The apriori algorithm is one of the algorithms that searches for frequent itemsets using the association rule technique (Fitri Marisa et al., 2021).

Broadly speaking, the way the apriori algorithm works is:

1. Formation of candidate itemset. Candidate k itemsets are formed from a combination of (k-1)-itemsets obtained from the previous iteration. One feature of the a priori algorithm is the removal of k-itemset candidates whose subsets containing k-1 itemsets do not belong to high-frequency patterns of length k-1.
2. Calculation of the support of each candidate k-itemset. The support of each candidate k-itemset is obtained by scanning the database to count the number of transactions that create all items in the candidate.

2.2.1 Association Rule / Association Rule

Association Rule is a data mining technique for finding association rules between combinations of items. Association Rule Mining is one of the existing techniques in data mining which aims to obtain association rules that can be obtained from various data sources, including those from transactional databases, data warehouses, and other information storage places (Ardilla et al., 2021). The a priori algorithm includes a type of association rule in data mining based on the fact that a priori uses data / prior knowledge of an itemset with a frequent frequency of occurrence or frequent itemset. Association rule mining finds interesting association or correlation relationships among a large set of data items ().

Association rule has the form LHS (left hand side)->RHS (right hand side) With the interpretation if every item in LHS will be purchased then in RHS will also be purchased (). Association rules are also often called association rules in the form of "if-then" or "if-then" is a bond or can also be called a "what to what" relationship. This rule is calculated from a set of data whose probability of the importance of an association can be known by two benchmarks, namely: support and confidence (Putra et al., 2023). Support is the percentage of combinations of these items in the database, while confiendence is the strength of the relationship between items in the association rule. Association rules not meeting the minimum support and minimum confiendence requirements will be discarded().

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

Apriori algorithm is a data linkage algorithm. Apriori is used to find associative rules or patterns between a combination of items.
Interestingness measures that can be used in data mining are: (Ardilla et al., 2021)

a. Support, is a measure that shows how much dominance an item or itemset has over the entire transaction.
b. Confidence, is a measure that shows the relationship between two items conditionally (based on a certain condition).

2.2.2 High Frequency Pattern Analysis

This stage looks for combinations of items that meet the minimum requirements of the support value in the database. The support value is the percentage of items or combinations of items in the database. The support value of an item is obtained by the following formula: (Fitri Marisa et al., 2021)

$$Support(A) = \frac{\sum Transaction\ containing\ value(A)}{\sum Transaction} \times 100\% \quad (1)$$

Meanwhile, the support value of 2-items is obtained from the following formula:

$$Support\ (\mathrm{A} \cup \mathrm{B}) = \frac{Number\ of\ Transactions\ containing\ A\ and\ B}{Total\ Transaction} \times 100\% \quad (2)$$

2.2.3 Association Rule Formation

After all the high-frequency patterns are found, then look for association rules that meet the minimum requirements for confidence by calculating the confidence of rule A U B. The confidence value of rule A U B is obtained by the following formula (Fitri Marisa et al., 2021):

$$Confidence = \mathrm{P(B|A)} = \frac{\text{Number of Transactions containing A and B}}{\text{Total Transaction containing A}} \times 100\% \quad (3)$$

2.3 Program Modeling

In this research, flowchart is used as a design stage from start to finish to make it easier to describe the program and pseducode as a language that almost resembles a programming language. Usually, pseudocode uses a language that is easily understood universally and is also more concise than algorithms.

2.3.1 Flowchart

Flowchart is a way of writing algorithms using graphical notation. Flowchart is a picture or chart that shows the sequence or steps of a program and the relationship between processes and their statements (Sulasmoro, 2022). Flowchart is a way to explain the stages of problem solving by presenting certain symbols that are easy to understand, easy to use and standard (Sulasmoro, 2022). The main purpose of using flowcharts is to simplify a series of processes or procedures in order to facilitate user understanding of the information. The following are symbols that can be used in describing flowcharts.

Table 1

Simbol Flowchart

No.	Symbol	Name	Explanation
1.		Communicator	Entry and exit symbols or as symbols connecting processes on the same sheet or page.
2.		Input/Output	Symbols to represent input and output processes.
3.		Process	Symbols used to represent a process
4.		Flow Line	Symbols used for shows the flow of a process

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

5.		Decision	Symbols used to select conditions in the program
6.		Defined Process	This symbol is used to indicate a process whose details are shown at another place
7.		Terminal Point	Symbols used for shows the beginning and end of a process
8.		Liaison	This symbol is used to indicate a section of the flow that is interrupted on another page
9.		Preparation	This symbol is used to give the initial value of a quantity

2.3.2 Pseudocode

Pseudocode is a way of writing algorithms that resemble high-level programming languages (Sa'adah et al., 2021). Usually, pseudocode uses a language that is universally understood and is also more concise than algorithms. Pseucodocede contains descriptions of computer programming algorithms that use simple structures from several programming languages but the language is only intended to be human readable.

From the pseudocode of the apriori algorithm, it explains the initial stages of the calculation. The Apriori algorithm is to display transaction data. Then look for 1-itemset support and so on based on data that has been input previously with the same minimum support. After obtaining the results of the n-itemset support, the confidence value is sought based on the results of eliminating the previous n-itemset support. After that, the confidence result is eliminated with the minimum confidence that has been determined previously. After doing all the calculations, we get the results of the association rules that we use as conclusions.

2.4 Application Support System

The application system is ready-made software that will be used to help carry out the work of its users. Examples such as word processing, database management system desktop publishing spreadsheets and so on.

2.4.1 Rapid miner

Rapid miner is a software used in data mining that is open source and also displays a GUI (Graphic User Interface) in its use (Jollyta et al., 2021). Rapid miner is also a solution for analyzing data mining. Rapid miner can also do text mining, namely analyzing text, extracting patterns and combining with statistical, database, and artificial intelligence methods. Rapid binary is also believed to be utilized to understand deep learning, machine learning and predictive analysis. Predictive analysis, one of which is rapid miner, can be used in the apriori algorithm where its use uses the association function in determining the final result.

Rapid miner uses various descriptive and description techniques in providing insight to users in making decisions or results.Rapid miner also uses java language so that it can work in all operating systems, Rapid miner has the following properties: (Sa'adah et al., 2021)

1. Written in java programming language so that it can be run on various operating systems
2. The knowledge search process is modeled as a tress operator
3. Built-in XML representation to ensure standardized format of data exchange
4. Scripting language allows for large-scale experiments and experiment automation.
5. Multi-layer concept to ensure efficient data display and ensure data subscription
6. Has a GUI, command line mode and java API that can be called from other programs.

2.4.2 Excel

Excel is an application in the form of a spreadsheet. Excel is one of the software in the Microsoft Office package. Excel is generally used for data processing in the form of numbers and calculations (Wibowo, 2023). Function and formula features or better known as excel formulas are formulas that are often used. Automatic data processing such as basic calculations, function usage, graph making, table processing and others (Kusrianto, 2019). In this research, excel is made as a database before working on the rapid miner.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

3. METHODS

3.1 Research framework

A research framework is the basic structure used to plan and organize a study. It provides direction and a framework for the researcher to identify the research problem, collect data, analyze the results, and conclude the findings. This is due to the fact that the subsequent research must be delivered in a sequential and on-track manner, hence the need for a research framework. Therefore, a research outline should be created before embarking on the research phase.

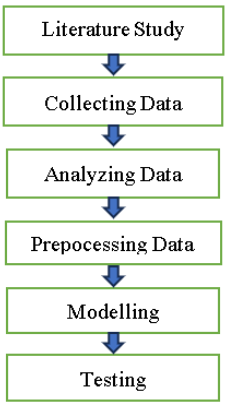

Fig. 2 Stages of the Research Framework

3.1.1 Literature Study

Literature study is carried out to collect knowledge by studying literature sources to be used as a reference for data related to the research to be carried out. At this stage researchers use literature sources in the form of books, journals and other scientific works.

3.1.2 Data Collection

The data collection techniques that will be used are:

1. Observation The researcher conducts a pre-research first to find problems that occur in determining the prediction of the availability of beauty products. From these problems will be formulated in this study so as to find what formulations need to be prepared in solving these problems. The data used by researchers is primary data derived from observations at cosmetic stores.
2. Library Research, namely by looking for references from library books and also from the internet, which are related to the object under study, where these references are used as a theoretical basis in completing this research.
3. Data collection is done by collecting historical data from October 2023 to January 2024, calculating the minimum support and minimum confidence values as benchmarks in the apriori algorithm process, the results obtained are outputs which are association rules that will be used by cosmetic stores.

3.1.3 Data Analysis

In this study, data analysis is needed so that it can be carried out properly. At this stage includes what needs are used in research, especially beauty product inventory data is the main focus. The data required is 500 data on the number of beauty products at Kazana Kosmetikeluti product names and available quantities.

In addition to beauty product inventory data, the Apriori algorithm was chosen because the Apriori algorithm is able to find patterns in beauty product sales data. These patterns indicate combinations of products that are often purchased together. This information is very useful for predicting which products are likely to be sold together in the future. By knowing the buying patterns, the Apriori algorithm can predict which products will sell out soon. This allows Kazana Kosmetik to order product stock on time and avoid stock-outs.

Table 2
Examples of data collected

No	Transaction Code	Selling price	Number of Sales
1	20240425221141	Skintific	274

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

2	20240425215622	Avoskin	269
3	20240425215246	Facetology	240
....	...	…	...
1400	20240425181147	Rollover Reaction	132

3.1.4 Data Preprocessing

After data collection and data analysis, the next stage is data preprocessing which is useful for removing inconsistent data so that clean data is obtained, data that will be used for clustering with the Apriori algorithm. At this stage, data selection and data transformation into numerical data are also carried out.

3.1.5 Modelling

Modelling is the process of planning or organizing structures, patterns, or plans needed to achieve certain goals. In this study, the design carried out is the design of the process described in the flowchart. The following is an apriori flowchart that aims to solve problems in beauty products in the following picture:

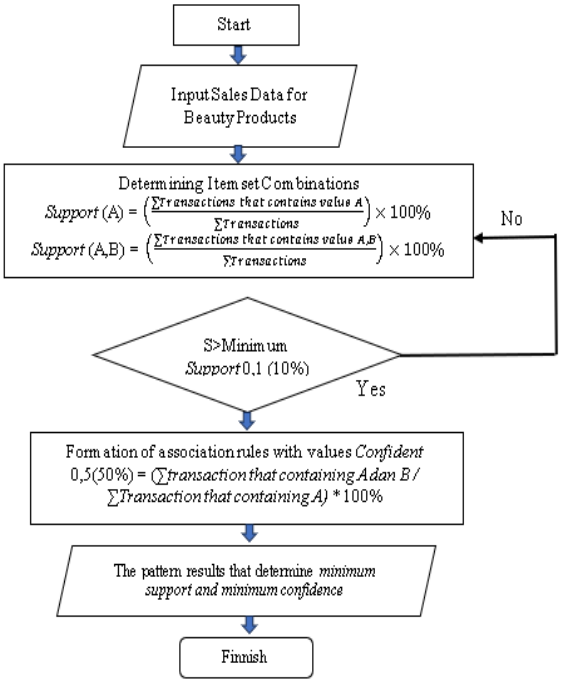

Fig. 3 Flowchart of Apriori Algorithm

The flowchart image of the a priori algorithm method above explains the calculation stages of the a priori algorithm from start to finish, which are as follows:

1. Input sales data of building tools and materials.
2. Determine candidate itemset combinations from beauty product sales data.
3. Determine the selected itemset combination.
4. If the support value is greater than the minimum support then the itemset combination will continue otherwise the combination search stops.
5. Then the formation of the association with the minimum confidence value that has been set before.
6. The result of the association rule pattern that finds frequent-itemset becomes the result of the association rule.

3.1.6 Testing

The testing stage is a stage to find out whether the rapid miner application that has been used is in accordance with its functions and outputs. Testing is carried out on availability predictions where the data inputted to the system

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

will be processed into itemsets, then itemsets are selected as needed and input minimum support in the fp-growth parameter and minimum confidence in the create association rules parameter after going through the calculation process, the output is obtained which is an association rule in the form of data, graphs and descriptions.

3.2 Discussion Plan

The discussion plan of this research is with the problem of beauty product inventory patterns at Kazana Kosmetik Medan city using rapidminer and excel applications so that the results of this research can later facilitate Kazana Kosmetik and get the results of the accuracy of the grouped data to become information data.

3.3 Research Location

This research was conducted at Kazana Kosmetik Store which is located at Jalan Perjuangan No.73C, Kel.Sei Kera Hilir I, Kec.Medan Perjuangan, Medan City, North Sumatra.

4. RESULT

4.1 Data Analysis

This research uses cosmetic sales data from Kazana Kosmetik Store collected for 2 months. The data used includes transaction IDs and a list of purchased products that include types of cosmetic products. In total, there were 1065 transactions analyzed to identify cosmetic purchase patterns in this store.

The first step in data analysis involved data cleaning and preparation. Transactions containing incomplete or irrelevant information were removed to ensure the quality of the data analyzed. For example, transactions with unidentified items or extremely unreasonable purchase amounts were excluded from the dataset. After the cleaning process, the remaining data is converted into a format suitable for use in the Apriori algorithm, i.e. a list of item-sets that records the products purchased in each transaction.

The Apriori algorithm was then applied to identify association patterns between cosmetic products that are frequently purchased together. This process involves several key steps, including the determination of minimum support and minimum confidence parameters. Minimum support is set to filter out the set of products that frequently co-occur in the dataset, while minimum confidence is used to evaluate the reliability of the association rules found. The Apriori algorithm searches for frequent itemsets or sets of items that frequently co-occur, and constructs association rules based on these frequent itemsets.

4.2 Data Representation

This research involves 1065 transaction data from total sales for 2 months at Kazana cosmetics. The data used in this research is data on products sold in 1 purchase. The following is a sample of the data used.

Table 3
Research Data Sample

Transaction	Products
1	['Primer', 'Powder', 'Lipstick', 'Eyeliner', 'Eyeshadow']
2	['Eyeshadow', 'Lipstick', 'Foundation', 'Primer', 'Blush']
3	['Powder', 'Blush', 'Eyeliner', 'Mascara', 'Eyeshadow']
…	…
10	['Foundation', 'Lipstick', 'Eyeshadow', 'Eyeliner']

4.3 Data Preprocessing

Data preprocessing is an important stage in this research to ensure that cosmetic sales data from Kazana Kosmetik Store is ready to be analyzed using the Apriori algorithm. This stage involves a series of steps that aim to clean, organize, and convert the raw data into a format suitable for further analysis. With a total of 1065 transactions analyzed, the following preprocessing steps were systematically applied.

a. Data Cleaning

The first step in preprocessing is data cleaning. This process involves removing transactions that are incomplete or contain irrelevant information. Transactions that have missing data or inconsistencies, such as unrecognized product IDs or unrealistic purchase amounts, are removed from the dataset. In addition,

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

duplicate data was also identified and removed to avoid bias in the analysis. The result of this cleaning process is a more accurate and representative dataset.

b. Data Transformation

After cleaning, the raw data is transformed into a more structured format. Each transaction is represented as a set of items that includes all the products purchased in that transaction. For example, if in one transaction there are purchases of products A, B, and C, the transaction is represented as {A, B, C}. This transformation is important to allow the Apriori algorithm to identify associations between products that are often purchased together. Here is how the data looks after going through the data transformation stage.

Table 4
Data Transformation Results

Transaction	Powder	Blush	Eyeliner	Eyeshadow	Foundation	Lipstick	Mascara	Primer
1	1	0	1	1	0	1	0	1
2	0	1	0	1	1	1	0	1
3	1	1	1	1	0	0	1	0
…	…	…	…	…	…	…	…	…
10	0	0	1	1	1	1	0	0

4.4 Application of Apriori Algorithm

The Apriori algorithm is used to identify association patterns between cosmetic products that are often purchased together. In its application, minimum support and minimum confidence parameters are determined to extract significant association rules. Minimum support is the minimum threshold for how often a group of products appear together in a dataset, while minimum confidence is the minimum threshold for how often an association rule is proven to be true in the context of an entire transaction. To use the Apriori algorithm and calculate the results manually, we will follow these steps:

a. Itemset Identification

Find the frequently occurring itemsets based on the specified minimum support threshold. Itemsets are combinations of products in a transaction. First, we need to list all the products. In this example we will only use sample data and set the support value = 0.2. In the Apriori algorithm, which is used to find patterns or rules in transaction data, the support value is one of the main metrics to measure how often an item or combination of items appears in the data. The higher the support value used, the more conclusions will be obtained, but it will reduce the accuracy of the conclusions.

Based on the data in table 4, the following product list is obtained.

1) Powder
2) Blush
3) Eyeliner
4) Eyeshadow
5) Foundation
6) Lipstick
7) Mascara
8) Primer

b. Form the frequent-1 item set and calculate its support.

Calculate the support for a 1-itemset using the following equation

$$Support(produk) = \frac{Jumlah\ Kemunculan\ Produk}{Total\ Transaksi} \quad (4)$$

By using this equation, the support value for 1-itemset is obtained as follows.

Support(Powder)=5/10=0.6
Support(Blush)=5/10=0.5
Support(Eyeliner)=5/10=0.5
Support(Eyeshadow)=5/10=0.5
Support(Foundation)=3/10=0.3
Support(Lipstick)=7/10=0.7
Support(Mascara)=5/10=0.5

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

Support(Primer)=6/10=0.6

By setting the support value = 0.2, all itemsets meet the threshold.

Formation of 2-itemset and calculate the support. Here is the support calculation for 2-itemset.

Support(Powder,Blush) = 3/10 = 0.3
Support(Powder,Eyeliner) = 3/10 = 0.3
Support(Powder,Eyeshadow) = 4/10 = 0.4
Support(Powder,Foundation) = 0/10 = 0.0
Support(Powder,Lipstick) = 4/10 = 0.4
Support(Powder,Mascara) = 2/10 = 0.2
Support(Powder,Primer) = 4/10 = 0.4
Support(Blush,Eyeliner) = 3/10 = 0.3
Support(Blush,Eyeshadow) = 3/10 = 0.3
Support(Blush,Foundation) = 2/10 = 0.2
Support(Blush,Lipstick) = 3/10 = 0.3
Support(Blush,Mascara) = 2/10 = 0.2
Support(Blush,Primer) = 2/10 = 0.2
Support(Eyeliner,Eyeshadow) = 3/10 = 0.3
Support(Eyeliner,Foundation) = 1/10 = 0.1
Support(Eyeliner,Lipstick) = 3/10 = 0.3
Support(Eyeliner,Mascara) = 2/10 = 0.2
Support(Eyeliner,Primer) = 2/10 = 0.2
Support(Eyeshadow,Foundation) = 2/10 = 0.2
Support(Eyeshadow,Lipstick) = 4/10 = 0.4
Support(Eyeshadow,Mascara) = 2/10 = 0.2
Support(Eyeshadow,Primer) = 3/10 = 0.3
Support(Foundation,Lipstick) = 2/10 = 0.2
Support(Foundation,Mascara) = 1/10 = 0.1
Support(Foundation,Primer) = 2/10 = 0.2
Support(Lipstick,Mascara) = 3/10 = 0.3
Support(Lipstick,Primer) = 4/10 = 0.4
Support(Mascara,Primer)=1/10= 0.2

With the determination of the support value = 0.2, the 2-itemset that meets the support value is determined so that 25 2-itemsets are obtained.

Table 5
Determination of 2-itemset

No.	Itemset	Support
1	{Powder, Blush}	0.3
2	{Powder, Eyeliner}	0.3
3	{Powder, Eyeshadow}	0.4
…	…	…
25	{Mascara, Primer}	0.2

Forming a 3-itemset and calculating its support

Here is the support calculation for 3-itemset.

Support({Powder,Eyeliner,Eyeshadow}) = 2/10 = 0.2
Support({Powder,Eyeliner,Lipstick}) = 2/10 = 0.2
Support({Powder,Eyeshadow,Lipstick}) = 2/10 = 0.2
Support({Powder,Eyeshadow,Primer}) = 2/10 = 0.2
Support({Blush,Eyeliner,Eyeshadow}) = 1/10 = 0.1
Support({Blush,Eyeshadow,Primer}) = 1/10 = 0.1

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

Support({Eyeliner,Eyeshadow,Primer}) = 1/10 = 0.1
Support({Eyeshadow,Lipstick,Primer})=1/10= 0.2

With the determination of the support value = 0.2, then determine the 3-itemset that meets the support value so that 5 3-itemsets are obtained.

Table 6
Determination of 3-itemset

No.	Itemset	Support
1	{Powder, Eyeliner, Eyeshadow}	0.2
2	{Powder, Eyeliner, Lipstick}	0.2
3	{Powder, Eyeshadow, Lipstick}	0.2
4	{Powder, Eyeshadow, Primer}	0.2
5	{Eyeshadow, Lipstick, Primer}	0.2

Form a 4-itemset and calculate its support.

The following is the support calculation for the 4-itemset.

Support({Powder,Eyeliner,Eyeshadow,Lipstick}) = 1/10 = 0.1
Support({Powder,Eyeliner,Eyeshadow,Primer}) = 1/10 = 0.1

Since none of the 4-itemset calculation results meet the support value = 0.2, then there is no 4-itemset obtained from the data used. From the calculation example above, it is concluded that by using 10 dataset samples, 29 purchase patterns are obtained as follows.

Table 7
Example of calculation results

No.	Itemset	Support
1	{Powder, Blush}	0.3
2	{Powder, Eyeliner}	0.3
3	{Powder, Eyeshadow}	0.4
…	…	…
29	{Powder, Eyeshadow, Lipstick}	0.2

The Apriori algorithm with a minimum support of 0.2 identifies itemsets that appear frequently in transactions. The results provide an overview of product combinations that are often purchased together by customers. By doing the calculation manually, we can clearly see how different product combinations have a significant frequency of occurrence.

The previous result is an example of applying the Apriori algorithm using 10 sample data and a support value = 0.2. For actual research, a support value of 0.7 is used because actual research requires more accurate results. By applying the apriori algorithm in the rapidminer application to 1065 data and the determination of the support value = 0.7, the final conclusion is obtained as follows.

```
Association Rules
[Primer] --> [Eyeshadow] (confidence: 0.406)
[Maskara] --> [Blush] (confidence: 0.408)
[Eyeliner] --> [Primer] (confidence: 0.413)
[Bedak] --> [Eyeshadow] (confidence: 0.413)
[Eyeshadow] --> [Primer] (confidence: 0.419)
[Foundation] --> [Blush] (confidence: 0.419)
[Eyeshadow] --> [Bedak] (confidence: 0.422)
```

Fig. 4
Association Rule Results

From the table above, it is known that there are 7 patterns that are most often done by buyers when buying

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

products at Kazana cosmetics. That is if you buy Primer then you will buy Eyeshadow, Mascara then you will buy Blush, Eyeliner then you will buy Primer, Powder then you will buy Eyeshadow, Eyeshadow then you will buy Primer, Foundation then you will buy Eyeshadow and Eyeshadow then you will buy Powder.

5. CONCLUSION

Based on the research that has been done, the following conclusions can be drawn: Apriori algorithm can be applied to sales transactions at Kazana Kosmetik to find purchase patterns. The results of applying the Apriori algorithm to 1065 transaction data at Kazana Kosmetik resulted in seven main buying patterns from buyers, with an average confidence value of 0.414. The purchasing patterns found are expected to be a reference for Kazana Kosmetik for product stock management. In addition, the findings from the Apriori algorithm can be used to design more effective marketing strategies, such as placing interconnected products in the same area of the store as well as offering bundling promotions to increase sales.

6. REFERENCES

Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Manuhutu, M. A., Ridwan, M., & Wardhani, A. K. (2021). *DATA MINING DAN APLIKASINYA*. Penerbit Widina. https://books.google.co.id/books?id=53FXEAAAQBAJ

Butar, A. M. B. (2022). *Penerapan Algoritma Apriori Pada Pengolahan Data Mining Untuk Mengetahui Pola Pembelian Konsumen Pd. Lucky Metal Part*. repositori.buddhidharma.ac.id. http://repositori.buddhidharma.ac.id/id/eprint/1484

Dari, W. (2022). Analisis Metode Apriori Untuk Memprediksi Persediaan Barang Pada Warung. *INSOLOGI: Jurnal Sains Dan Teknologi*, *1*(4), 438–447. https://doi.org/10.55123/insologi.v1i4.807

Fitri Marisa, S. K. M. P., Anastasia Lidya Maukar, S. T. M. S. M. M. T., & Dr. Tubagus Mohammad Akhriza, S. S. M. M. S. I. (2021). *Data Mining Konsep Dan Penerapannya*. Deepublish. https://books.google.co.id/books?id=BtlVEAAAQBAJ

Harahap, F., Saragih, N. E., & ... (2022). Implementasi Data Mining dalam Memprediksi Stok Herbal menggunakan Algoritma Apriori. *JURNAL MEDIA …*. http://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/3937

Jollyta, D., Siddik, M., Mawengkang, H., & Efendi, S. (2021). *Teknik Evaluasi Cluster Solusi Menggunakan Python Dan Rapidminer*. Deepublish. https://books.google.co.id/books?id=3rcgEAAAQBAJ

Kusrianto, A. (2019). *Menguasai Formula dan Fungsi Excel 2007-2019*. Elex Media Komputindo. https://books.google.co.id/books?id=0tKdDwAAQBAJ

Megawati, M. (2022). *Penerapan Data Mining Dalam Analisis Aturan Association Rule Berdasarkan Data Transaksi Penjualan Menggunakan Metode Apriori dan Fp-Growth Pada Anugrah …*. repository.unilak.ac.id. https://repository.unilak.ac.id/id/eprint/2887

Pratiwi, D., & Wibowo, J. S. (2023). Implementasi Algoritme Apriori Pada Sistem Persediaan Obat Apotik Puskesmas. *… Teknik Informatika Dan Sistem …*. http://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/article/view/1106

Putra, R. F., Zebua, R. S. Y., Budiman, B., Rahayu, P. W., Bangsa, M. T. A., Zulfadhilah, M., Choirina, P., Wahyudi, F., Andiyan, A., & Efitra, E. (2023). *DATA MINING : Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia. https://books.google.co.id/books?id=zLHGEAAAQBAJ

Rahmatullah, S., Mintoro, S., & Permatasari, K. (2022). PENERAPAN METODE ALGORITMA APRIORI DALAM MEMPREDIKSI PENJUALAN SPAREPART MOTOR (PT. LAUTAN TEDUH INTERNIAGA DEALER YAMAHA KOTABUMI). *Jurnal Informasi Dan Komputer*, *10*(2).

Sa'adah, U., Rochayani, M. Y., Lestari, D. W., & Lusia, D. A. (2021). *Kupas Tuntas Algoritma Data Mining dan Implementasinya Menggunakan R*. Universitas Brawijaya Press. https://books.google.co.id/books?id=SI1TEAAAQBAJ

Salmon, S., Azahari, A., & Yusnita, A. (2022). Prediksi Persediaan Bahan Baku Makanan Menerapkan Algoritma Apriori Data Mining. *Building of Informatics …*. http://ejurnal.seminar-id.com/index.php/bits/article/view/2563

Setiawan, Z., Fajar, M., Priyatno, A. M., Putri, A. Y. P., Aryuni, M., Yuliyanti, S., Widiputra, H., Meilani, B. D., Ibrahim, R. N., & Azdy, R. A. (2023). *BUKU AJAR DATA MINING*. PT. Sonpedia Publishing Indonesia. https://books.google.co.id/books?id=1nLVEAAAQBAJ

Sulasmoro, A. H. (2022). *Buku ajar algoritma dan pemrograman I*. Penerbit P4I.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4259

Submitted : 9 July 2024
Accepted : 20 July 2024
Published : 27 July 2024

https://books.google.co.id/books?id=gUNjEAAAQBAJ

Ulkhairi, M. (2020). *Implementasi Data Mining Penjualan Kosmetik Dengan Algoritma Apriori*. repository.upbatam.ac.id. http://repository.upbatam.ac.id/id/eprint/82

Urva, G., Albanna, I., Sungkar, M. S., Gunawan, I. M. A. O., Adhicandra, I., Ramadhan, S., Rahardian, R. L., & Handayanto, R. T. (2023). *PENERAPAN DATA MINING DI BERBAGAI BIDANG : Konsep, Metode, dan Studi Kasus*. PT. Sonpedia Publishing Indonesia. https://books.google.co.id/books?id=uq6-EAAAQBAJ

Wibowo, H. S. (2023). *Tutorial Excel : Panduan Lengkap untuk Pemula*. Tiram Media. https://books.google.co.id/books?id=2IS2EAAAQBAJ

* Corresponding author