**IMPLEMENTING APRIORI ALGORITHM IN SUPERMARKET SALES DATA**

1st Jovinka Aphellia Salva  
*Computer Science Study Program,   
University of North Sumatra*Medan, Indonesia   
jovinkav@gmail.com2nd Najwa Afifi Situmorang  
*Computer Science Study Program,   
University of North Sumatra*Medan, Indonesia   
najwaafifi121@gmail.com

3rd Dewi Sartika Br Ginting S.Kom., M.Kom  
*Computer Science Study Program,   
University of North Sumatra*Medan, Indonesia   
dewidintingdg90@gmail.com

***Abstract*** - Using association rule mining techniques, we aim to distribute groups of commonly occurring items and generate meaningful association rules that can provide valuable insights into customer purchasing behavior.

The dataset consists of further details such as invoice ID, product line, and payment model and has been pre-processed to simplify the main process. Our analysis reveals that the support value to Fashion accessories (0.178) which is the highest support value among others.

***Keywords-association, a priori, sales, data***

# **INTRODUCTION**

A priori algorithm is the process of extracting information from a database and often generates elements or sets of elements and candidates to form association rule mining to obtain minimum support and minimum confidence values.

For a large enough database, the a priori algorithm will generate a large number of frequently used item/itemset patterns, as it needs to create candidates and keep track of recurring databases.

In this study, we applied the Aprior algorithm to a supermarket store dataset containing supermarket transaction sales data. The main objective of this research is to find common product groups that consumers often buy together. By identifying these buying behaviors, supermarket owners can gain insight into their marketing tactics and sales techniques to improve overall performance.

Before applying the Aprior algorithm, we perform preliminary data mining to understand the characteristics of the dataset and identify important information such as best-selling products, most popular payment methods, and sales by gender and month. The purpose of this step is to gain insights into customer buying patterns and prepare the data for further analysis.

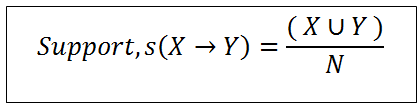
# **THEORY**

Apriori is an algorithm for mining target sets and learning association rules iteratively through a relational database. The process involves identifying specific objects that are found repeatedly in the database and then gradually increasing the number of such objects until they are available for analysis**.**

Mining frequently used products and their association rules is achieved through the use of the Apriori algorithm. In general, the apriori algorithm works on databases that contain a large number of transactions. For example, consumer goods, but at the Grand Bazaar.

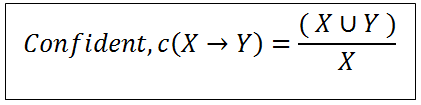
Support, confidence, and lift are the three main components in the association data mining process using the apriori algorithm.

The percentage form of the number of occurrences for a particular combination of items is called the Support(s) value.



The importance of support values in association rules is emphasized, as low support levels indicate that associations are rare in the data set (all event data).

Calculating the percentage of the accuracy of the association rules that will be generated is called the Confident(c) value.



The magnitude of Y is defined as high confidence for events containing X.

***Lift Ratio*** is a parameter that measures the strength of association rules created by support values and beliefs.

# **RESEARCH METHODOLOGY**

This research applies the a priori algorithm from Data Mining to extract data related to supermarket sales.

The data used are supermarket sales records that contain information about all sales transactions, such as products purchased, payment methods and customer demographics.

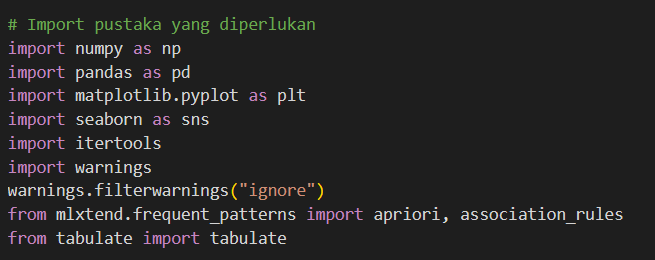
## Data and Preprocessing

* 1. Data Source:

The dataset used comes from a Kaggle CSV file titled "supermarket\_sales.csv". The data in this dataset includes various attributes such as transaction ID, city, member, gender, product purchased, price, quantity, date, payment method, and others.

* 1. Data Reading:

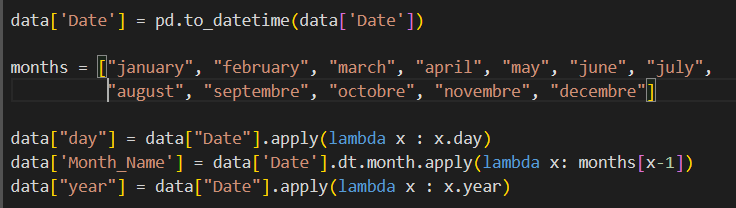
The dataset was read using the `pandas`, `numpy`, `matplotlib.pyplot`, `seaborn`, `itertools` and `warnings` libraries with the following commands:



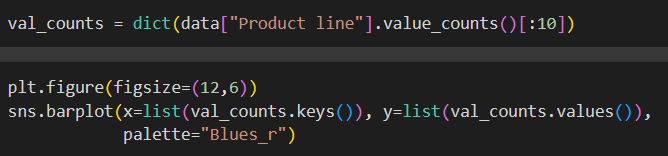
* 1. Data Preprocessing:

The data was processed to prepare it for further analysis.

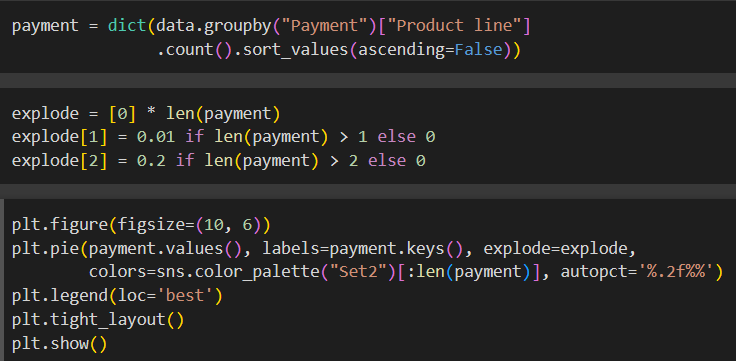
* Converts a date column to datetime type.
* Added a column that separates the day, month, and year information from the date column.
* Delete columns that are not required for analysis.



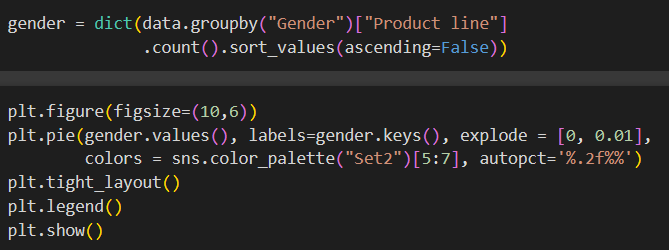
1. *Exploratory Data Analysis* 
   1. Visualization of Number of Products Purchased



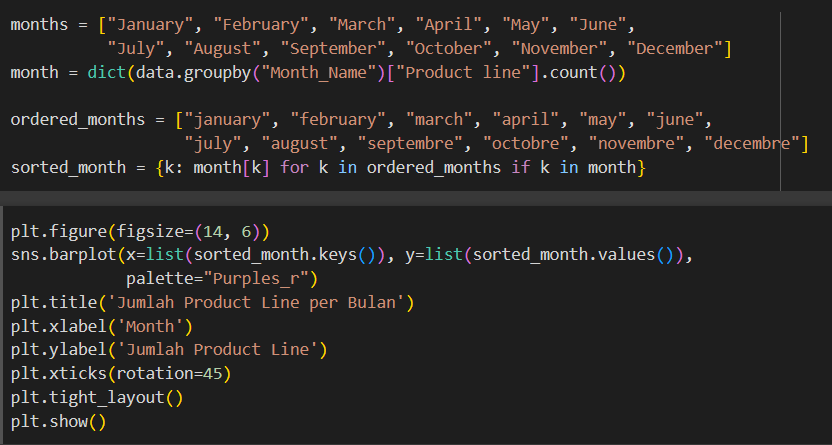
* 1. Visualization of Payment Method Features



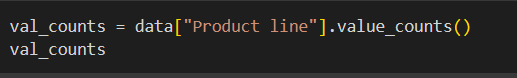
* 1. Gender Feature Visualization



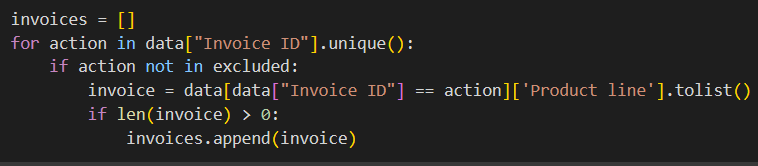
* 1. Moon Feature Visualization



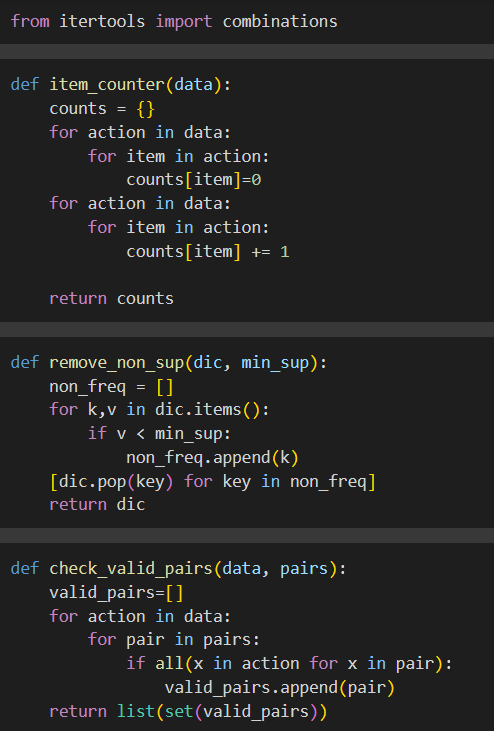
1. *Data Preprocess (for Pattern Data Mining)*
   1. Summation of Each Purchased Product

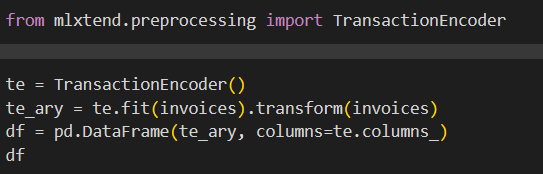


* 1. Transaction ID Appearance



1. *Implementation of Apriori Algorithm*



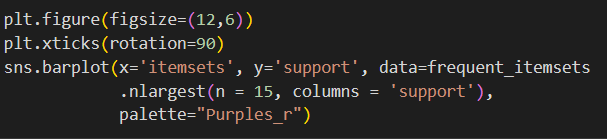
1. *Retrieving the Final Result and Comparing it with the Original Algorithm/Facts*
   1. Display TransactionEncorder
   2. Formation of Frequent Itemsets

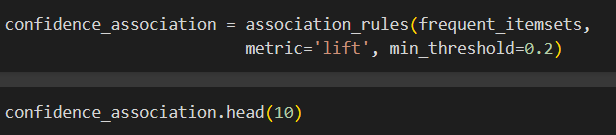
Frequent itemsets are formed using the `apriori` function from the `mlxtend.frequent\_patterns` library with a predefined minimum support.

* 1. Formation of Largest Value in Frequent Itemsets



* 1. Image Visualization of Frequent Itemsets



1. Interpretation of Association Rules

## Units

# **RESULTS AND DISCUSSION**

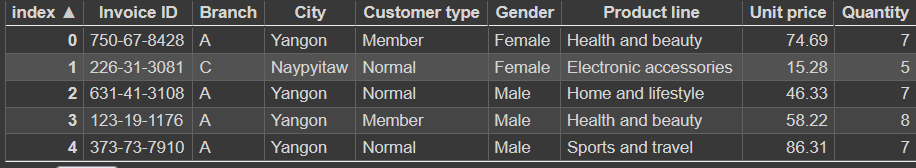


Figure 1: Learning Data has been Retrieved and Run

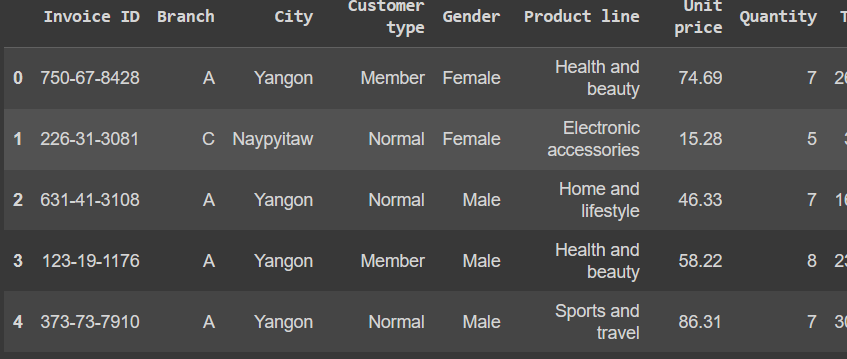


Figure 2: Data Preprocessing

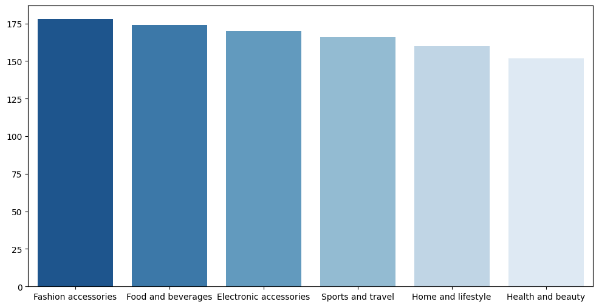


Figure 3: Visualization of Number of Products Purchased

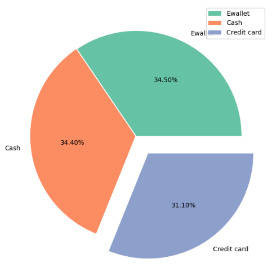
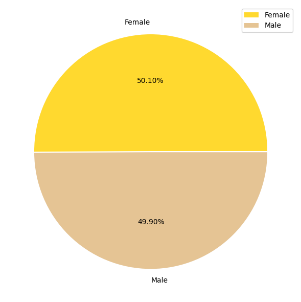


Figure 4: Payment Method Fitue Visualization



\\

Figure 5: Gender Fitue Visualization

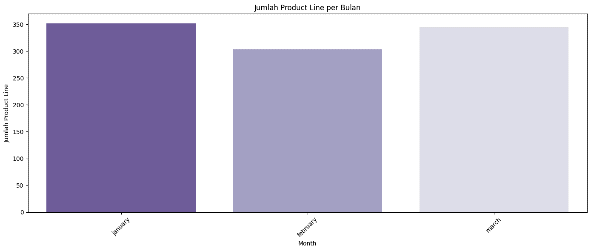


Figure 6: Moon Feature Visualization

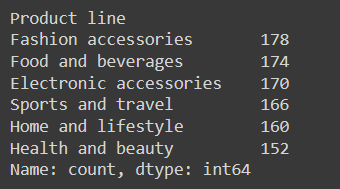


Figure 7: Number of Each Product Purchased



Figure 8: Ercorder Transaction Table

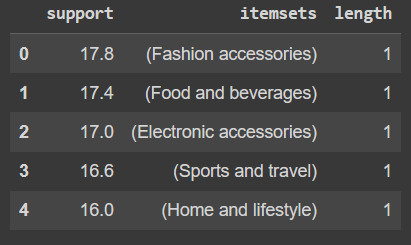


Figure 9: 1 Itemset Support Table



GGGGGGGG

Figure 10: Support Table

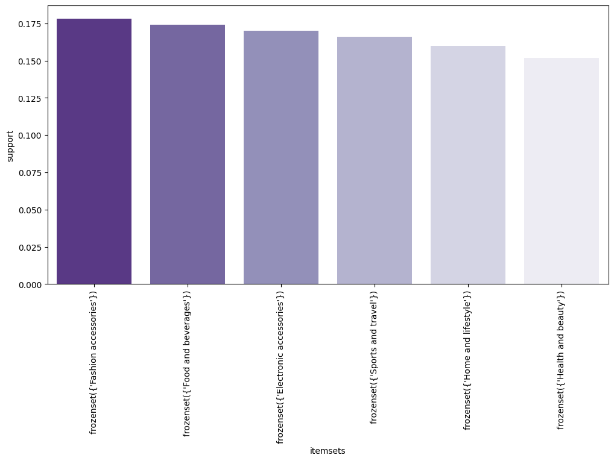


Figure 11: Figure drawing of Support

# **CONCLUSIONS**

We use the Apriori algorithm to apply recurring pattern mining analysis to event sales supermarket data in this study. Functions such as "item counter", "*remove\_non\_sup*", "*check\_valid\_pairs",* "*pair\_counter", "unique\_elements*", and "apriori" were created to implement the Apriori algorithm manually.

This analysis shows common clusters or groups of products that are frequently purchased by supermarket customers in large quantities in different time periods (e.g. 2, 3, 4, 5 items out of stock). Item sets with item lengths of 2 and 3 represent the most frequent purchase patterns observed in the transaction data.

This research not only looks for the number of frequently viewed items, but also calculates the trust value between specific items or product groups. Trust is calculated using the equity\_on\_items and equity\_on\_sets functions. These indicate how often a product or product line is purchased when another product or product line is purchased in the same transaction.

The results of model mining and continuous trust analysis help supermarkets understand consumer purchasing behavior. This information can be used in marketing strategies, campaigns, product placement or new product development to meet customer preferences. Overall, this research shows that Aprior algorithm can find shopping patterns that are useful for business decision making through pattern mining analysis on supermarket transaction data.

# **LITERATURE**

[**https://en-m-wikipedia-org.translate.goog/wiki/Apriori\_algorithm?\_x\_tr\_sl=en&\_x\_tr\_tl=id&\_x\_tr\_hl=id&\_x\_tr\_pto=tc**](https://en-m-wikipedia-org.translate.goog/wiki/Apriori_algorithm?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc)

[**https://www-javatpoint-com.translate.goog/apriori-algorithm?\_x\_tr\_sl=en&\_x\_tr\_tl=id&\_x\_tr\_hl=id&\_x\_tr\_pto=tc**](https://www-javatpoint-com.translate.goog/apriori-algorithm?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc)

[**https://colab.research.google.com/drive/13bAvG56tL1kY113duD9pCNr1s1qXFFUB?authuser=0#scrollTo=tnSj\_c8nI311**](https://colab.research.google.com/drive/13bAvG56tL1kY113duD9pCNr1s1qXFFUB?authuser=0#scrollTo=tnSj_c8nI311)