Determining Sales Patterns of Beauty Products Using the Apriori Algorithm in Data Mining

Windi Maharani 1*, Raissa Amanda Putri 1

*Correspondence Author: e-mail: windimaharani937@gmail.com

1 Computer Science Study Program; Universitas Islam Negeri Sumatera Utara; Address: Jl. Lap. Golf, Kp. Tengah. Kec. Pancur Batu, Kab. Deli Serdang, Sumatera Utara 20353, Medan, Indonesia; e-mail: windimaharani937@gmail.com e-mail: raissa.ap@uinsu.ac.id

Abstract

The use of data mining technology is a must to extract information that can help in making better decisions, especially in the beauty product industry, where competition is getting tougher. Changes in lifestyle and an increase in people's income lead to an increase in public consumption of beauty products. From these problems, a data mining system was created to determine the sales pattern of beauty products using the Apriori algorithm method. The goal is to get information about products that are often purchased by consumers so that they can prepare their needs with business strategies. The use of Apriori can perform the process of extracting information from the database to find association rules between a set of data or a combination of items/itemsets. The results of this study formed an application that was built using the python programming language to analyze product sales at Underprice Skincare stores using the apriori algorithm. The data generated from the application process of applying data mining will be used as a reference for stores to prepare stock availability and can design the right business strategy to increase sales of beauty products at Underprice Skincare stores.

Keywords: Data mining, Beauty products, Apriori, Python, Sales.

1. Introduction

Data has become one of the most valuable assets for businesses in the rapidly growing digital era. The opportunity to gain valuable insights is huge when a lot of data collected from various business transactions are stored. The use of data mining technology has become a necessity to extract information that can help in making better decisions, especially in the beauty product industry, where competition is getting tougher (Munthe & Juledi, 2021)(Putra et al., 2023).

Beauty products refer to a wide variety of items and formulations designed to enhance or maintain one's physical appearance (Maryani et al., 2022). These different types of products include skincare, hair, nails, perfumes, cosmetics, and other items used to improve skin health and aesthetics. At present, beauty products have developed in such a way along with changes in lifestyle, increased public income, and higher education levels. Changes in lifestyle and an increase in people's income lead to an increase in people's consumption of beauty products.

Data mining is a process of sifting through large data sets to identify patterns and relationships that can help solve business problems through data analysis (Agustiani et al., 2020)(Tan et al., 2019). Data mining is a technique used to derive valuable information from a sizable data set through various processes (Muhammad Arhami & Muhammad Nasir, 2020)(Fitri Marisa et al., 2021). This technique helps characters analyze the sales data patterns stored in the character database and process the contents of sales transaction data (Mulaab, 2021).
One of data mining method is association rule mining. It is a data mining method for finding association rules between combinations of items. Association techniques involve developing models and analyzing customer behavior when purchasing goods (Haryandi et al., 2021).

Based on these problems, it is found that by applying data mining techniques, it can identify patterns that may be difficult to detect manually, allowing companies to make more informational and evidence-based decisions. One data mining method that has proven effective in association analysis is the apriori algorithm (Sianturi et al., 2019) (Maryani et al., 2022).

One of the data mining algorithms, the a priori algorithm, performs the process of extracting information from a database to find association rules between a set of data or a combination of items/itemsets (Umar et al., 2022). Therefore, the Apriori algorithm will be suitable to be applied to solve these problems.

Previous research entitled "Implementation of Data Mining for Sales of Cosmetic Products at PT Natural Nusantara Using the Apriori Algorithm" by (Wardani & Kristiana, 2020) argues that the use of the a priori algorithm can help management to place items that are commonly purchased by consumers, making it easier for consumers to be able to buy these items.

The purpose of this research is to analyze and apply data mining with the apriori algorithm method to obtain information about products that are often purchased by consumers so that they can prepare needs with business strategies.

2. Research Method

A research framework is a structure or conceptual framework used to design, develop and structure a study. A research framework provides the researcher with a theoretical foundation for designing the research methodology, analyzing the data, and interpreting the findings. It assists the researcher in framing relevant research questions, selecting appropriate methods, and structuring an analytical framework that fits the research objectives. The research method that the author uses is a quantitative research method which uses numerical data and emphasizes the research process on measuring objective results using statistical analysis. The research framework in analyzing sales patterns of beauty products using the a priori algorithm method. The research framework will be presented in Figure 1 below:

![Research Framework Diagram](source: Research Result (2024))

Figure 1. Research Framework
Data collection techniques are methods or strategies used to collect information or data relevant to the objectives of a particular research or study. This technique helps researchers to obtain the data needed to answer research questions or achieve research objectives in a systematic and directed manner. The techniques used in this research are interviews, surveys, documentation.

Based on the results of observations and interviews at the Underprice.skincare store located at Jl. Jamin Ginting No.656, Titi Rantai, Kec.Medan Baru, Medan City, North Sumatra, primary data from the sales pattern of beauty products during November 1, 2023 - December 31, 2023 as many as 1000 transactions are obtained as follows:

<table>
<thead>
<tr>
<th>Number of Transactions</th>
<th>Date</th>
<th>Item Name</th>
<th>Sold Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01-Nov-23</td>
<td>Emina Magic Potion Lip Tint - Summer</td>
<td>25</td>
</tr>
<tr>
<td>1</td>
<td>01-Nov-23</td>
<td>Emina Poppin Matte Special Edition - 04 Snatched</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>01-Nov-23</td>
<td>Wardah Colorfit Last All Day Lip Paint - 15 Dazzling New York</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>01-Nov-23</td>
<td>Emina Daily Matte Loose Powder - 02 Natural Beige</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>01-Nov-23</td>
<td>Wardah Everyday Luminous Two Way Cake - 01 Light Beige</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>01-Nov-23</td>
<td>Wardah Acnederm Series - Day Moisturizer</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>01-Nov-23</td>
<td>Wardah Everyday BB Cream - Natural 15ml</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>02-Nov-23</td>
<td>Emina Bright Stuff Night Cream</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>31-Des-23</td>
<td>Wardah Colorfit Perfect Glow Cushion - 32N Neutral Beige</td>
<td>2</td>
</tr>
</tbody>
</table>

Source: Research Result (2024)

From table 1 there are 1000 transactions at the Underprice.skincare store which is transaction data, the following is the result of 1-itemset from the transaction data above:

<table>
<thead>
<tr>
<th>No</th>
<th>Item Name</th>
<th>Sold Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Emina Avocado Honey Face Mask 60ml</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td>Emina Bare With Me Mineral Mild Foundation - 01 Light</td>
<td>108</td>
</tr>
<tr>
<td>3</td>
<td>Emina Bright Stuff Face Serum 30ml</td>
<td>53</td>
</tr>
<tr>
<td>112</td>
<td>Wardah Colorfit Fresh Lip Ink Serum - 02 Smart Cookie</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>7773</td>
</tr>
</tbody>
</table>

Source: Research Result (2024)

Flowchart is a section with certain symbols that describe the sequence of a process in detail and the relationship between a process and other processes in a program. The following is an a priori flowchart that aims to solve problems in beauty products in the following image:
3. Results and Analysis

3.1 Discussion

In this study, we will analyze the sales patterns of beauty products found at the Underprice Skincare store located at Jl. Jamin Ginting No. 656, Titi Rantai, Kec. Medan Baru, Medan City, North Sumatra. The analysis process is carried out by data mining using the Apriori algorithm. Furthermore, the data generated from the data mining process using the Apriori algorithm will be used as a reference for stores to prepare stock availability and can design the right business strategy to be able to increase sales of beauty products at Underprice Skincare stores.

3.2 Data Analysis

The data used in this study is in the form of sales data at the Underprice Skincare store starting from November 1, 2023 to December 31, 2023. For more details, sales data at the Underprice Skincare store can be seen in table 3.
3.3 Apriori Algorithm Calculation

In the apriori algorithm calculation process, the data mining technique that will be used is the association technique. The apriori algorithm calculation process has the aim of finding association rules, which association rules will later be used as a benchmark to see several combinations of drug items and medical devices that are most often purchased by consumers. This can make it easier for Underprice Skincare stores to make decisions such as making stock planning and implementing the right business strategy in increasing sales. The steps to create a data mining model using the apriori algorithm are as follows: Determine what data you want to process, determine the minimum support and minimum confidence values, compile the association rules formed.

In this study, the process of finding association rules was carried out with the provisions of a minimum support value of 0.01 and a minimum confidence of 0.5. Where the data to be processed is the sold product data contained in table 4.1. Furthermore, the amount of support will be sought from the transaction data. The trick is to divide between the number of item occurrences and the number of all transactions.

\[
\text{Support (Emina Magic Potion Lip Tint} - \text{Summer}) = \frac{36}{4011} = 0.008975
\]

The value of 36 is obtained from the number of transactions that have the item "Emina Magic Potion Lip Tint - Summer" in it, while the value of 4011 is obtained from the total number of transactions based on the data contained in table 3. By using this calculation process for each product data sold, each Association Rule value for product sales at the Underprice Skincare store is as follows:

<table>
<thead>
<tr>
<th>No.</th>
<th>Antecedents</th>
<th>Consequents</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>'Emina Aqua Infused Sleeping Mask', 'Emina Cheek Lit Blush Apple Crumble 3.5g', 'Emina Cheek Lit Blush Apple Crumble 3.5g', 'Wardah White Secret Pure Treatment Essence 8ml'</td>
<td>'Emina Cheek Lit Blush Apple Crumble 3.5g', 'Emina Aqua Infused Sleeping Mask'</td>
<td>0.0163831</td>
<td>0.5531914</td>
<td>33.7659574</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>'Emina Aqua Infused Sleeping Mask'</td>
<td>'Emina Aqua Infused Sleeping Mask'</td>
<td>0.0163831</td>
<td>1</td>
<td>33.7659574</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>'Emina Aqua Infused Sleeping Mask'</td>
<td>'Emina Aqua Infused Sleeping Mask'</td>
<td>0.0132325</td>
<td>0.1909090</td>
<td>6.44622824</td>
<td>7</td>
</tr>
<tr>
<td>74</td>
<td>'Wardah Lightening Two Way Cake - 06 Light Tan 12.5g'</td>
<td>'Wardah Hairfall Treatment Shampoo'</td>
<td>0.0132325</td>
<td>0.6363636</td>
<td>12.7836593</td>
<td>8</td>
</tr>
</tbody>
</table>

Source: Research Result (2024)
From table 4, there are several column headings that must be understood first, which will be explained as follows:

a. Antecedents means that the item is the cause of other items being sold.
b. Consequents means the sale of item B which will be sold if item A is sold.
c. Support shows how often the items are bought together in one transaction.
d. Confidence is the level of trust (in percentage) if buying item A will also buy Item B.
e. Lift is the relationship of the rules that form a positive relationship (there is a correlation), negative, or no correlation.

From Table 4, it can be seen that 74 rules that pass to form an association, one of which is if the customer buys 'Emina Aqua Infused Sleeping Mask', then he buys 'Emina Cheek Lit Blush Apple Crumble 3.5g' which has a Support value of 0.016 which means there are 1.6% of the number of transactions that have the product 'Emina Cheek Lit Blush Apple Crumble 3.5g' in it. Confidence of 0.553 means that there are a number of transactions consisting of the product item 'Emina Aqua Infused Sleeping Mask' has a 5.3% chance of also buying 'Emina Cheek Lit Blush Apple Crumble 3.5g' in the same transaction. The entirety of the resulting association rules can be seen in table 5.

<table>
<thead>
<tr>
<th>No.</th>
<th>Rule</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>If customer bought 'Wardah Colorfit Ultralight Matte Lipstick - 05 Blossom Pink', then he/she bought 'Wardah Colorfit Perfect Glow Cushion Refill - 32 Natural Beige'</td>
<td>0.013</td>
<td>0.840</td>
</tr>
<tr>
<td>2</td>
<td>If the customer bought 'Wardah Colorfit Perfect Glow Cushion Refill - 32 Natural Beige', then he/she bought 'Wardah Colorfit Ultralight Matte Lipstick - 05 Blossom Pink'</td>
<td>0.013</td>
<td>0.875</td>
</tr>
<tr>
<td>3</td>
<td>If the customer bought 'Emina Cheek Lit Blush Apple Crumble 3.5g', then he/she bought 'Emina Aqua Infused Sleeping Mask'</td>
<td>0.016</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>If a customer bought 'Emina Magic Potion Lip Tint - Summer', then she bought 'Wardah Nature Daily Seaweed Cleansing Micellar Water 240ml'</td>
<td>0.011</td>
<td>0.472</td>
</tr>
<tr>
<td>74</td>
<td>If a customer bought 'Wardah Lightening Fesh BB Tint - 01 Neutral Fair 15ml', then she bought 'Emina Bright Stuff Face Toner - 50ml'</td>
<td>0.011</td>
<td>0.246</td>
</tr>
</tbody>
</table>

Source: Research Result (2024)

The design of the Data Mining Implementation application in Analyzing Beauty Product Sales Patterns Using the Apriori Algorithm Method was built using Visual Studio Code software with the Python programming language can be seen as follows:

3.4 Application Test

Tests were conducted to see the results of data mining using the apriori algorithm on sales data contained in the Underprice Skincare store. The test results can be seen as follows:
1. Sales dataset selection process

Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Kecantikan Menggunakan Algoritma Apriori

Source: Research Result (2024)

Figure 3. Dataset File Selection

3.5 Application Design

After selecting the sales dataset, the application will display the contents of the dataset in the table as follows:

Source: Research Result (2024)

Figure 4. Displaying the Dataset into a Table
Furthermore, users can see the best-selling products based on the data that has been inputted as follows:

Source: Research Result (2024)

Figure 5. Displaying Best-Selling Product Data

2. Association Calculation Process

After selecting a dataset, the user can then start the analysis process using the Apriori algorithm to get association rules from product sales as follows:

Source: Research Result (2024)

Figure 6. Displaying the Association Rule
Furthermore, users can see the results of stock provision recommendations based on calculations that have been generated as follows:

- **Figure 7. Displaying Stock Item Recommendations**

  **Rekomendasi stok barang untuk dibeli (contribution):**

  1. ‘Emina Aqua Infused Sleeping Mask’ (0.0138965210453554)
  2. ‘Wardah UV Shield Active Serum SPF 50 25ml’ (0.017002247300839427)
  3. ‘Wardah Everyday BB Cream - Natural 15ml’ (0.01589495600068546)
  4. ‘Emina Daily Matte Compact Powder - 04 Mocca’ (0.0259683644496033)
  5. ‘Emina Bright Stuff Face Serum T.5ml’ (0.0232359417783762)
  6. ‘Wardah White Secret Pure Treatment Essence 8ml’ (0.012323324014552146)
  7. ‘Wardah Colorfit Ultra Light Matte Lipstick -05 Blossom Pink’ (0.01158449865432041)
  8. ‘Wardah Instaperfect Liquid Lipstick Set - 05’ (0.0142004400243026)
  9. ‘Wardah Colorfit Last All Day Lip Paint - 16 Dazzling New York’ (0.0113121434893927)
  10. ‘Wardah Colorfit Perfect Glow Cushion Refill - 32 Natural Beige’ (0.01115131308532762)
  11. ‘Wardah Colorfit Last All Day Lip Paint - 07 Brick On’ (0.012389891014712964)
  12. ‘Emina Watercolor Lip Serum - 01 Dawn’ (0.012919354715299)

  Source: Research Result (2024)

On this page, users can also directly see the correlation between products in sales based on the data that has been selected as follows:

- **Figure 8** is the final result of the analysis of the mining process using the apriori algorithm. Where the data will display the relationship between products in sales. For example, there is a relationship if a customer buys 'Wardah Colorfit Perfect Glow Cushion Refill - 32 Natural Beige', then he buys 'Wardah Colorfit Ultralight Matte Lipstick - 05 Blossom Pink' with a confidence value of 0.875 or 87.5%. So that in the future these data can be used by Underprice Skincare stores in stocking and designing the right business strategy to get better company profits.
4. Conclusion

Based on the results and discussions that have been produced in this study, it can be concluded that in this study a system was produced that can analyze the sales data of beauty products contained in the Underprice Skincare store. The data processed using the data mining process in this study is sales data at the Underprice Skincare store from November 1, 2023 to December 31, 2023. The data mining process on sales data is carried out using the Apriori algorithm. The resulting system can display product sales patterns at Underprice Skincare stores. The system can display the best-selling products based on the dataset that has been selected. The system can display beauty products that are recommended for the provision of stock items. The system in this study was built using the Python programming language. The data tested is sales data at the Underprice Skincare store that has been processed and stored as a file in .csv format.

Acknowledgements

During the process of preparing this final project, the author received a lot of direction, guidance, support and prayers from various parties. To my parents who have educated and provided support and invaluable love and prayers, namely my beloved father Sukardi and my beloved Mamak Supriani. The author would also like to thank: Prof. Dr. Nurhayati, M. Ag as the Rector of the State Islamic University of North Sumatra, Dr. Zulham, S.H.I., M. Hum as the Dean of the Faculty of Science and Technology of the State Islamic University of North Sumatra, Mr. Ilka Zufria, M.Kom as the head of the Computer Science Study Program of the State Islamic University of North Sumatra, Dr.
Determining Sales Patterns of Beauty Products Using the Apriori Algorithm in Data Mining

M. Fakhriza, ST., M.Kom as the Secretary of the Computer Science Study Program of the State Islamic University of North Sumatra Medan, Mrs. Raisa Amanda Putri S.Kom., M.TI as my supervisor who has provided a lot of direction, guidance and support to me to complete this Final Project, Mr. Dr. Ali Darta, M.A as my academic supervisor who has provided guidance, advice and motivation to me, All lecturers, staff and employees of the Computer Science Study Program, Faculty of Science and Technology, State Islamic University of North Sumatra Medan, To brother Hari Pourwanto Ritonga as a resource person who is willing to help me for research at Underprice skincare Medan city, To my brother Dendy Rahmadi who always motivates and encourages in working on this thesis and is always the best for his sister, To Muhammad Alwi Fahrozi Hasibuan thank you for being a loyal supporter in life. Every advice, every support is a motivation and thank you also for being a home where the place to vent about this final project which has a lot of drama, to my friend Yessi Mayasari thank you for being a part of compiling this thesis which always wants to be a place to be bothered about these dramas, which always accompanies the running of this final project, to F2 housing thank you for being a good family, Especially Yana Wulandari who has been enthusiastic in working on this final project who has been willing to help provide motivation and a place to complain in this final project, The extended family of the author's late grandfather Musni and Grandma Ginem who always provide enthusiasm, support and prayers, Friends in arms especially the Department of Computer Science class of 2020, And finally to myself. Windi Maharani. Thank you for surviving this far. Thank you for still choosing to try and celebrate yourself to this point, even though you often feel desperate for what you are trying and have not succeeded, but thank you for remaining a human being who always wants to try and is not tired of trying, no matter how difficult this final project is, you have completed it as well and as much as possible. Finally, the author hopes that this final project is useful for readers and can be used to increase knowledge for all those who need it. Thank you.

Author Contributions
Windi Maharani proposing the topic, designing related to the Implementation of Data Mining to Determine Sales Patterns of Beauty Products Using the Apriori Algorithm (Case Study: Underprice.Skincare Medan).

Conflicts of Interest
The authors declare no conflict of interest.

References

PIKSEL status is accredited by the Directorate General of Research Strengthening and Development No. 225/E/KPT/2022 with Indonesian Scientific Index (SINTA) journal-level of S3, starting from Volume 10 (1) 2022 to Volume 14 (2) 2026.


