International Journal of Advanced Multidisciplinary(IJAM)
E-ISSN : 2829-6192
P-ISSN : 2829-6184
https://greenpub.org/IJAM greenation.info@gmail.com +62 812-1046-7572
DOI: https://doi.org/10.38035/ijam.
Received: xx June 202x, **Revised:** xx June 202x, **Publish:** xx July 202x

Application of the Apriori Algorithm in PT XYZ

Cipriano Bruno das Neves[1], Ridwan[2]
[1]Universidade Oriental Timor Lorosa'e, Dili, Timor Leste, ciprianoneves85@gmail.com
[2]Institut Teknologi dan Bisnis Dewantara, Bogor, Indonesia, ridwans70@gmail.com

Corresponding Author: ciprianoneves85@gmail.com[1]

Abstract: Competition in the business world, especially in the increasingly difficult printing world, requires developers to find strategies to increase orders for printed products ordered. An increasing number of order data every day can be used to develop marketing strategies if processed correctly. A priori algorithms include the type of association rules in data mining. One of the stages of association analysis that attracts many researchers to produce efficient algorithms is the analysis of high-frequency patterns (frequent pattern mining). The importance of an association can be known by two benchmarks, namely: support and confidence. Support (support value) is the percentage of the combination of these items in the database, while confidence (certainty value) is the strength of the relationship between items in association rules.
Keyword: apriori, data mining, and ordering

INTRODUCTION

The quick advancement of computer science and technology has made it essential and vital, particularly for the growth of management and organizations in all types of businesses and companies, public and private. Due to intense rivalry in the business world, particularly in the printing sector, developers must devise a plan that will boost product orders at printing companies. The length of the data will increase daily with the printing activity. Consequently, every business needs to have a robust data processing system in place in order for the information gathered from these transactions to be utilized for creating monthly or yearly reports. The company can use and process the data in addition to using it as an archive (John W. Rittinghouse & James F. Ransome, 2009).

Printing is a technology or art form that swiftly duplicates images, such as words or drawings, on various surfaces like paper and fabric. Because science and technology are developing at a faster rate, the printing industry has evolved into a more sophisticated and modern sector (Sugiharjo et al., 2024). Indonesian printing businesses—large, medium, and small—are expanding quickly. This boom also occurred in the city of Bekasi, which was home to a sizable number of printing enterprises. The printing industry is experiencing heightened business competitiveness as a result of the growth in the number of printing enterprises (Cyril Stanley Smith, 2023).

In order to stay ahead in this competitive market, printing businesses in Bekasi are constantly investing in new technologies and equipment to improve their efficiency and quality of work. Many companies are also focusing on expanding their services to include digital printing, large format printing, and customized printing solutions to meet the diverse needs of their clients. As a result, the printing industry in Bekasi is not only growing rapidly but also becoming more diverse and innovative. With the right strategies and investments, printing businesses in Bekasi are poised to continue thriving in this dynamic market (Burger et al., n.d.).

METHOD

Data Mining

Data mining is the process of automatically analyzing and extracting knowledge using one or more machine learning algorithms. Induction-based learning, which is the process of developing broad idea definitions through the observation of particular instances of concepts to be learned, is one of the other definitions. Applying scientific techniques to data mining is known as Knowledge Discovery in Databases (KDD). Data mining is one step in the KDD process in this context (Kerdprasop & Kerdprasop, 2012).

The process of filtering extremely large amounts of data that are stored in storage and applying pattern recognition techniques, such as statistical and mathematical methods, to identify meaningful relationships, patterns, and new trends is known as data mining (Pondri, 2023).

Data mining involves extracting and analyzing large sets of data to uncover hidden patterns and relationships that can provide valuable insights for decision-making. By utilizing advanced algorithms and machine learning techniques, data mining can help organizations identify trends, predict future outcomes, and optimize their operations. It plays a crucial role in various industries, including finance, healthcare, marketing, and more, by enabling businesses to make data-driven decisions and gain a competitive advantage in the market (Liane Collone, 2013).

Overall, data mining is a powerful tool that allows businesses to harness the vast amounts of data available to them and turn it into actionable insights. By understanding consumer behavior, market trends, and internal operations more effectively, organizations can adapt and innovate to stay ahead of the competition. With the continuous advancements in technology and data analytics, data mining will continue to play a vital role in shaping the future of business decision-making (Craig Vauhan, 2013).

High Frequency Patterns with Apriori Algorithms (C.L. Philip Chen, n.d.).

This step searches the database for item combinations that satisfy the minimal requirements of the support value. The following formula is used to determine the support value of an item:

$$support(A) = \frac{jThe\ amount\ of\ transaction\ containedA}{total\ transactions}$$

Meanwhile, the support value of 2 items is obtained using the formula:

$$suppotr(A,B) = P(A \cap B)$$

$$support(A,B) = \frac{\sum transaction\ contains\ A\ and\ B}{Total\ transactions}$$

Items with a frequency of occurrence greater than the given minimum value (∅) are displayed in the frequent itemset. All itemsets that appear more than or twice at the same moment, for instance, are said to be common when ∅ = 2. Fk represents the set of frequent k-item sets.

Association Rules

The formula for confidence is calculated by dividing the number of transactions that contain both items A and B by the number of transactions that contain item A. This calculation helps organizations determine the strength of the relationship between two items and make informed decisions based on the data. By leveraging data mining techniques to analyze consumer behavior and market trends, businesses can gain valuable insights that drive strategic decisions and ultimately lead to success in the competitive market landscape (Putra & Hendriyani, 2023).

By assessing the confidence of associative rules A → B, association rules that satisfy the minimal requirements for confidence are identified once all high frequency patterns have been identified. The following formula yields the confidence value of the rules A → B:

$$confidence = P(B / A) = \frac{\sum transactioninA\ and\ B}{\sum transaction\ A}$$

Support × Confidence must be ordered in order to choose the association rules. The number of rules that yield the best results is equal to n.

RESULTS AND DISCUSSION

Data requirements

The data used in the implementation of a priori algorithm are outlined in the following pattern:

Table 1. Frequency Patterns

No.	*Itemset*
1	Folder, Letterhead, Brochures
2	Folder, Brochures, Letterhead
3	Folder, Brochures, Letterhead
4	Brochures, Folder, Name Cards
5	Brochures, Folder, Name Cards
6	Folder, Brochures, Letterhead
7	Brochures, Letterhead, Invitations
8	Brochures, Folder, Envelopes
9	Brochures, Folder, Letterhead
10	Brochures, Invitations, Envelopes

From table 1 above, tabular results are generated in the following table:

Table 2. Tabular Transaction data

No	*A*	*B*	*C*	*D*	*E*	*F*
1	1	0	1	1	0	0
2	1	0	1	1	0	0
3	1	0	1	1	0	0
4	1	0	0	1	1	0
5	1	0	0	1	1	0
6	1	0	1	1	0	0
7	0	1	1	1	0	0
8	1	0	0	1	0	1
9	1	0	1	1	0	0
10	0	1	0	1	0	1

Formation of the itemset

Process of forming C1 or referred to as 1 itemset with a minimum amount of support = 55% With the following formula:

$$Support(A) = \frac{\sum transaction\ contains\ A}{\sum transaction} * 100\%$$

The following is the calculation of the formation of 1 itemset:

$$= \frac{\sum transactions\ contain\ Cake\ Boxes}{\sum 12} = \frac{10}{12} * 100\% = 83{,}33\%$$

$$= \frac{\sum transaction\ contains\ Invitation}{\sum 12} = \frac{2}{12} * 100\% = 16{,}67\%$$

$$Support\ (A) = \frac{\sum transaction\ contain\ A}{\sum transaction} \text{ x } 100\%$$

$$= \frac{\sum transaction\ contain\ Folder}{\sum 10} = \frac{8}{10} \text{x} 100\% = 80\%$$

$$= \frac{\sum transaction\ contain\ Letterhead}{\sum 10} = \frac{6}{10} \text{x} 100\% = 60\%$$

$$= \frac{\sum transaction\ contain\ Brochures}{\sum 10} = \frac{10}{10} \text{x} 100\% = 100\%$$

$$= \frac{\sum transaction\ contain\ Envelope}{\sum 10} = \frac{2}{10} \text{x} 100\% = 20\%$$

$$= \frac{\sum transaction\ contain\ Name\ Card}{\sum 10} = \frac{2}{10} \text{x} 100\% = 20\%$$

$$= \frac{\sum transaction\ contain\ Invitations}{\sum 10} = \frac{2}{10} \text{x} 100\% = 20\%$$

Based on the description above, it can be made in table 3

Table 3. Support of each item

Itemset	*Support*
Folder	80%
Invitation	20%
Letterhead	80%
Brochure	100%
Name card	20%
Envelope	20%

Combination 2 itemset

The process of forming C2 or referred to as 2 itemset with a minimum amount of support = 55%, The following is the calculation of the formation of C2 or 2 itemset:

$$= \frac{\sum transactions\ contain\ Folder\ and\ Invitations}{\sum 10} * 100\%$$
$$= \frac{0}{10} * 100\% = 0$$
$$= \frac{\sum transaksi\ mengandung\ \text{Folder}\ dan\ \text{Letterhead}}{\sum 10} * 100\%$$
$$= \frac{5}{10} * 100\% = 50\%$$
$$= \frac{\sum transactions\ contain\ \text{Folder}\ and\ Brochures}{\sum 10} * 100\%$$
$$= \frac{8}{10} * 100\% = 80\%$$
$$= \frac{\sum transactions\ contain\ \text{Folder}\ and\ \text{Name Card}}{\sum 10} * 100\%$$
$$= \frac{2}{10} * 100\% = 20\%$$

$$\frac{transactions\ contain\ \text{Folder}\ and\ \text{Envelope}}{\sum 10} * 100\%$$
$$= \frac{1}{10} * 100\% = 10\%$$
$$= \frac{\sum Transaction\ contains\ Invitation\ and\ Letterhead}{\sum 10} * 100\%$$
$$= \frac{1}{10} * 100\% = 10\%$$

Table 4 2-itemset candidates

Itemset	***value***	***Support***
Folder, Invitation	0	0
Folder, Letterhead	5	50%
Folder, Brochure	8	80%
Folder, Name Card	2	20%
Folder, Envelope	1	10%
Invitation, Letterhead	1	10%
Invitation, Brochure	2	20%
Invitation, Envelope	1	1%
Letterhead, Brochure	6	60%
Letterhead, Envelopes	0	0
Brochures, Envelopes	2	20%

The minimum support set is 55%, so the combination of 2 items set that does not meet the minimum supportwill be removed, looks like table 5 below:

Table 5 Minimum Support 2 itemset 55%

Itemset	***Support***
Folder, Letterhead	50%
Folder, Brochure	80%
Letterhead, Brochure	60%

The final association rules are ordered based on the minimum support and minimum confidence that has been determined, can be seen in table 6 below:

Table 6 Final Association Rules

Aturan	Support	Confidence
If ordering a Folder, it will order a Brochure	55%	90%

So, based on table 6 above, the most ordered items are Folder and Brochures. With the knowledge of the most ordered items, the company can find out the most ordered items.

CONCLUSION

The apriori approach can be used to find the most ordered printed materials in the given data by examining products that satisfy the least support and minimum confidence. Brochures and folders are the most often ordered commodities; nevertheless, processing large amounts of data might cause difficulties in calculating support and setup.

One way to address this challenge is to use data preprocessing techniques to clean and organize the data before applying the apriori algorithm. This can involve removing duplicate entries, handling missing values, and standardizing data formats to ensure accurate results. Additionally, businesses can consider implementing automated tools or software solutions to streamline the data mining process and improve efficiency. By investing in the right technology and expertise, organizations can harness the power of data mining to make informed decisions and stay ahead of the competition (Panggabean et al., 2023).

However, it is important to note that data preprocessing is not always foolproof and can sometimes introduce bias or errors into the dataset if not done correctly. For example, removing duplicate entries may inadvertently delete important information, and standardizing data formats could potentially distort the original meaning of the data. Therefore, careful consideration and validation are necessary to ensure the quality and reliability of the results obtained from data mining techniques.

While data preprocessing can introduce errors, the benefits of harnessing data mining for informed decision-making outweigh the risks if done correctly. Proper validation and oversight can help mitigate bias and ensure accurate results.

REFERENCE

Burger, N., Chazali, C., Gaduh, A., Rothenberg, A. D., Tjandraningsih, I., & Weilant, S. (n.d.). *Reforming Policies For Small And Medium-Sized Enterprises In Indonesia*. www.tnp2k.go.id

C.L. Philip Chen. (n.d.). *Data-intensive applications, challenges, techniques and technologies: A survey on Big Data*.

Craig Vauhan. (2013). *Data Science for Business*.

Cyril Stanley Smith. (2023). *Art, Technology, and Science: Notes on Their Historical Interaction*.

John W. Rittinghouse, & James F. Ransome. (2009). *Cloud Computing Implementation, Management, and Security*.

Kerdprasop, K., & Kerdprasop, N. (2012). Bridging Data Mining Model to the Automated Knowledge Base of Biomedical Informatics. In *International Journal of Bio-Science and Bio-Technology* (Vol. 4, Issue 1).

Liane Collone. (2013). *A Taxonomy of Classification in Data Mining*.

Panggabean, H. L., Firdiansyah Suryawan, R., Sani, I., & Sjarifudin, D. (2023). *Artificial Intelligence (AI) in Support of Marketing*. *1*(2). https://doi.org/10.38035/dit.v1i2

Pondri, F. (2023). *Analysis Repair End Voltage and shrink Technical with Reconfiguration Network 20 KV distribution on Feeder PT PLN (Persero) Service Unit Silago Customer (ULP) Sitiung*. *1*(1). https://doi.org/10.38035/dit.v1i1

Putra, R. M., & Hendriyani, Y. (2023). *Ordering Information System Design Weddings Organizer Putri Minang uses YII2 Framework*. *1*(1). https://doi.org/10.38035/dit.v1i1

Sugiharjo, R. J., Niken Purbasari, R., Rahmat, A., & Paijan, P. (2024). The Effect of Compensation on the Performance of Automotive Company Sales Division Employees: The Role of Work Discipline as a Mediator. *DIJMS*, *5*(6). https://doi.org/10.31933/dijms.v5i6