

Designing a mobile based-application using Apriori algorithm: A Business development strategy during pandemic COVID-19

Anastasia Aurelia^{1*}, Ni Putu Leona Yorita Dewi², Willy Tanupratalla³, Zafira Tsaqifah⁴, Lina Lina¹

¹Information Systems Department, Faculty of Information Technology, Universitas Tarumanagara

²Pharmacy Department, Faculty of Pharmacy, Universitas Mahasaraswati Denpasar

³Information Technics Department, Faculty of Information Technology, Universitas Tarumanagara.

⁴Digital Business Department, Faculty of Economic and Business, Institut Informatika dan Bisnis Darma Jaya

*anastasia.825180062@stu.untar.ac.id

ABSTRACT

It is very crucial for retailers to improve their marketing and customer relationship management strategies. Hence, this paper will cover the topic about how to utilize the association rule method to analyse the customers' purchasing behaviour and patterns based on market basket analysis. Association rule method can find the relationships between large data set of secondary files from several e-commerce platforms. This method can be performed by utilizing the cutting-edge data innovation and programming applications like data mining and Apriori algorithm libraries of Python programming language in Anaconda Navigator. By applying this Apriori algorithm, the transaction data obtained in case studies becomes easier to manage.

Keywords: E-commerce, Data mining, Market Basket Analysis, Apriori Algorithm, Association Rules

1. INTRODUCTION

The COVID-19 virus, SARS-CoV-2 causes major death since that infectious disease spreads easily. The ability of COVID-19 virus to sustain on surfaces makes it more contagious and more challenging to curb [1]. Countries worldwide tried to bring down the expansion of the virus by restriction of international travel, prohibition of large-scale gatherings, business closures, and mandatory quarantines that has evolved in huge effect to the global economy [2]. The effect of the COVID-19 crisis in the global stock market was the most obvious consequence. Global stock market decreased and lost \$6 trillion in over six days from 23 to 28 February 2020. Closures, highest and lowest stock prices, and general economic activity

were affected by the rising number of lockdown days, international travel restrictions, and monetary policy decisions. [3]. The COVID-19 pandemic impacts negatively on physical activity. The level of physical activity around the world has decreased during the social distancing policy on COVID-19 pandemic [4]. Pandemic issues have led people to look for other ways to interact with others due to the social distancing and nationwide lockdown's policy [5]. The use of digital technology drastically increased after the initiation of pandemic lockdown. These phenomena have a close correlation with the major increase on the amount of time spent on gadgets online as people's way to interact with others during the social distancing [6]. With

these phenomena, digital technology can be an effective way to do digital marketing and promotion for business owners which aim to increase their income and economy [7].

Communication and Technology helps businesses develop, create connections, strengthen the business, and allow people to learn from each other. Communication in the shape of marketing can value the brand or product to customers for the purpose of selling and promoting the product [7]. Medium of marketing is mostly through website, application, and social media. Along with that, the most efficient way to develop a campaign is with digital tools, so it is easier for the content to be available for the public and reach the audience using the search engine techniques [8]. With search engines, people who are wishing to find information about the product with the keyword that resembles what they need [9].

The phrase "Data mining" bring up to the process of extracting information from a database. There are many techniques in data mining to choose, one of them is apriori. Apriori algorithm is data retrieval algorithms using association rules to discover the association or relation between item combinations [10]. Using apriori algorithms to data mining techniques increase the efficiency and the speed on the process of combining pattern of itemset combination. In the implementation, apriori algorithm can be used to find the common item set run by another set of data. Therefore, apriori algorithms show recommendations for some related product to the consumers and develop the marketing of the related product [11].

2. LITERATURE REVIEW

Data mining is computational operation of pattern discovery in a large data that involves artificial intelligence, statistics, machine learning, and database systems. Data mining is used to transform the extracted information from a dataset to an understandable form [13]. Association rule is a method to discover

relation between variables in a database mining. This information will give knowledge to the business owner in observing the behavior of the customers [14].

Apriori algorithm is a classic algorithm that is used to explain current behavior, predicting future outcomes and to provide support for decision making and other business purposes [14]. Apriori arranged to run on databases with transaction history such as item combinations purchased by customers to discover the relationship among the items using mining association rules [13]. Apriori algorithm implementation can be used to determine tourism visit patterns in Bali. By knowing the pattern of tourism visit, all relevant stakeholders can prepare everything and maintain the quality for the tourists [11]. Resolution making and judging consumer spending behavior is a big challenge for an organization to keep going on their position in the market competition [12]. In business case, apriori algorithm is used to determine customers buying pattern with great precision and it can impact the daily sales along with the business development [13].

In Safitri and Purnomo's research [15], the apriori algorithm used to find the attachments between each products that are purchased by customers and the result of it would be useful to plan business's strategies, such as advertising and promoting. The data structure used the "No_SO" format, which is the identifier to determine association rules. Apriori algorithm produces product recommendations based on sales transactions. According to Alfianzah, Handayani, and Myrniyati [16] Affinity analysis or market basket analysis is a rule that states the association between items. Association rule can be determined by three parameters, namely:

- a. Supports (support value) : The level of the blend of things in the data set.
- b. Confidence (certainly value) : The strength of the connection between things in an affiliation rule.

- c. Lift : Number that incorporates the ensuing isolated by complete number of exchanges (certainty partitioned by assumption).

3. METHOD

3.1. Market Basket Analysis

The technique utilized is the investigation and examination in Indonesia dependent on understanding social data about prime clients' buying designs across classes, focusing on fragment explicit strategically pitching and up-selling drives. Market Basket Analysis combined with analytical techniques that advanced in retail marketing would grant understanding in today's Internet and digitally authprized customer [17]. Market Basket Analysis (MBA) shows the information about super-market pricing plan of action. MBA helps in enlarging strategies such as in profitatble advertising, promotion targets and revenue growth stimulant [18]. MBA is data mining methodes called frequent set purchasing pattern where it focal point on the finding of purchase patterns by extracting the associations from the transaction data of the store. The MBA defines the item set that is purchased together and helps in the designing and improving the promotion strategy based on the product's purchase. It also helps in judging and estimating consumer behaviour [19].

3.2. Data mining

Data mining is an act that includes collecting and using historical data to receive the patterns or relations in data pool. The ending result from data mining can be used to make hereafter decision accurately. The development of data mining has caused the use of pattern recognition to decrease because it has become a part of data mining [20], then, it will be processed into a display that displays the results of the data mining process in regions of Indonesia. Structured data is uploaded to a cloud server where various machine learning algorithms and data mining method is used for information and knowledge extraction [21]. The Business Data Analytic Framework for the

application model can be seen in Figure 1. Showing Internet of Things (IoT), device that can sense data, communicate with each other and potentially actuate using devices such as sensors, beacons, cameras, Wireless Networks and actuators. IoT usually connect in a huge number, and it can help retrieve valuable data that can be analyzed and used to optimize the supply chain, staff deployment, storage management and finally respond with actuation [22].

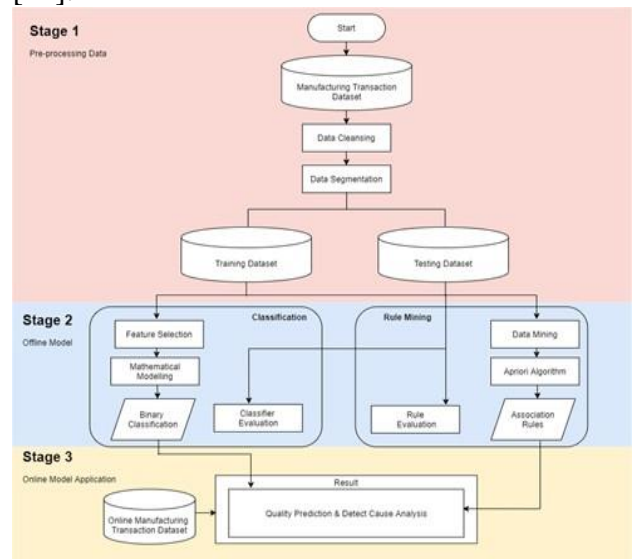


Figure 1. Business Data Analytic Framework for the application model

3.3. Mining Association Rules

Association rule is recognized with the statement and information of "what item has relationship with other". The information might be in a form of report on exchange action by the clients at the store. From that data, there has a solid association with the examination of client trade information data set to determine the tendency for a set of bought item together, hence, affiliation rule is as often as possible alluded to as market bin examination [23],[24]. To choose the rules of the Association, it is a must to determine the assist anfaith to limit whether the resulting rules are attractive or not [25]. MBA is a data mining technique called frequent set purchase pattern where it focal point on finding the purchase patterns by extracting the associations from the transaction

data of the store. It informs the item set that is purchased together and helps in the designing and improving the promotion campaigns based on the purchase of products. It also supports in judging and sensing consumer behaviour [26]. Market basket analysis (MBA) is also known as affinity analysis and associations rule mining. Various algorithms such as Apriori etc. are available for MBA performance [21]. The methodology to analyze the association is divided into 2 part, as follows :

3.3.1. Analysis of the highest frequency pattern.

Mining association rules are data mining techniques to find and analyze the connection between a combination of items. It is one of the most important method in data mining [27]. Whether an association rule is important or not can be determined by two parameters, the support (value of supporting / support) and confidence (value certainty) [21][28][29]. The calculation of support can be done with equations 1 and 2 as shown below:

$$Support(P) = \frac{\sum (Number\ of\ Transaction\ containing\ P)}{\sum Transaction}$$

While the support values for the 2 items are obtained from the formula as shown below:

$$Support(P, Q) = \frac{\sum (Number\ of\ Transaction\ containing\ P\ \&\ Q)}{\sum Transaction}$$

3.3.2. Rules of association formation.

After the Apriori algorithm can discover high-frequency patterns that have a frequency above a certain threshold in a database. The iterative process will run until the algorithm is unable to find a new itemset and produces an association rule, with an "if-then" pattern based on minimum confidence and minimum support [29]. The proportion of transactions with item P, in which item Q also appears divided by all transactions that contain Z. Confidence

calculation can be done with the following equation as shown below:

$$Confidence = Z(P|Q) = \frac{\sum (Transaction\ containing\ P\ \&\ Q)}{\sum (The\ Total\ Processing\ P)}$$

3.4. The lift ratio

The power possessed by the rules of association result is shown by the lift ratio. All transactions are measured by how likely item P is purchased when item Q is purchased, expressed as (P → Q), The Lift Ratio have a job on analyzing the strength of associations between products [29]. The formula of the Lift Ratio can be seen in equation 4 as shown below:

$$Lift\ Ratio = \frac{P \rightarrow Q}{Confidence(P \cap Q)} = \frac{Confidence(P \cap Q)}{Support(P)Support(Q)}$$

3.5. Apriori algorithm

Apriori algorithm is one of the most powerful calculation methods that is used for mining the data itemset and the use for affiliation rules on the conditional data sets. The affiliation rules are managed by apriori algorithm and the common pattern in the retail location of the data set. From that data, there has a strong association with the assessment of customer trade data informational index to decide the penchant for a purchased thing with what thing, henceforth, connection rule is as regularly as

conceivable insinuated as market canister assessment. Items are the elements that differentiate connection between. A gathering of items is called a 'set' [30].

$$X = \{x_1, x_2, x_3, \dots, x_n\}$$

Exchanges are instances of things matching together. For instance, an online bookshop, a trade might be the gathering of articles read in a singular visit to the webpage. We have a thing set for every exchange at that point [30].

$$y_n = \{p_a, p_b, \dots, p_k\}$$

Rules are the articulations of the structure

$$\{x_1, x_2, \dots\} \rightarrow \{x_n\}$$

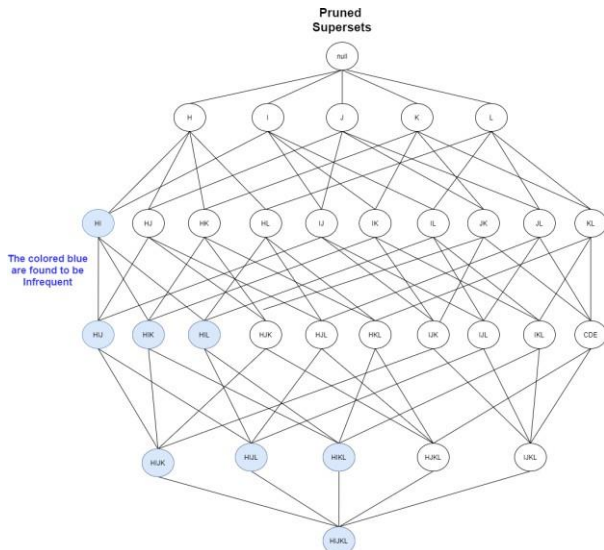


Figure 2. Apriori Algorithm

This is a bottom-up approach where the calculation begins from every single item in the itemset list. Then, contender are generated by self-joining. One at a time, the itemsets will extend the length. The subset test implementation is used at each stage and the itemsets that consist of the infrequent subsets are pruned. The calculation methods will repeat in action until no more itemsets can be derived from the data.

4. RESULTS AND DISCUSSION

The result of this research is an application that can be accessed by buyers and sellers with the aim of providing an effective way for analysing data to assist in judging behavioural information result about the pattern of customers purchases across categories. Users can see an overview and predictions regarding the types of categories and strategies by viewing it as a principles to improve workflow productivity and increasing employee engagement to help innovative technologies and e-retailing companies in enhancing customer loyalty during and after the pandemic. The proposed application can be

seen in Figure 3. where the user is asked to enter an email and password before using the application, which then allows the user to use features such as observing the type of categories that have been selected and analysing the data.

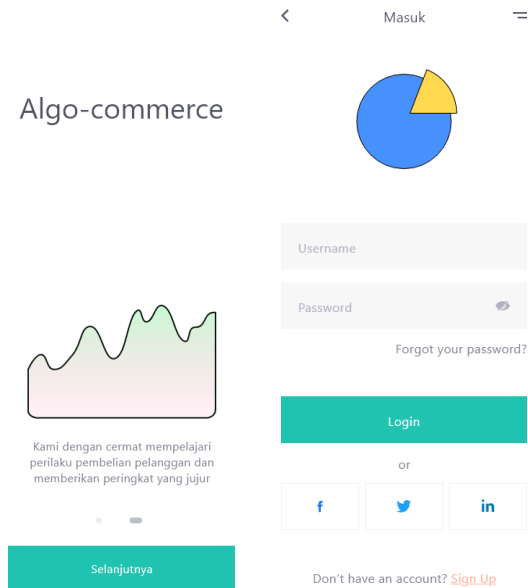


Figure 3. App Introduction and Login display

The application has features for customers selecting across E-commerce and categories which can be seen in Figure 4.

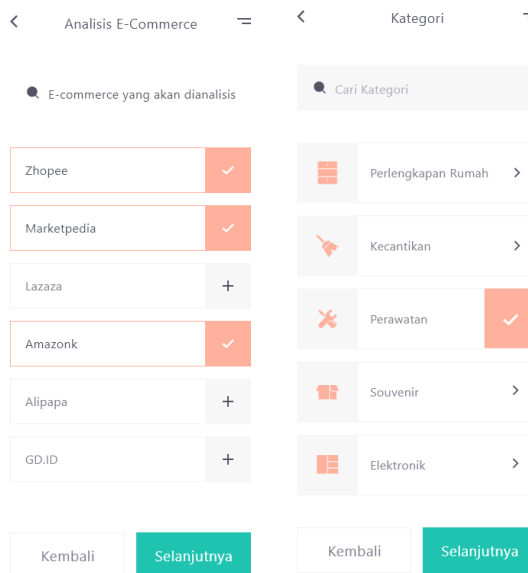


Figure 4. App selecting across category

The feature for overview and predictions regarding the types of categories is shown in Figure 5

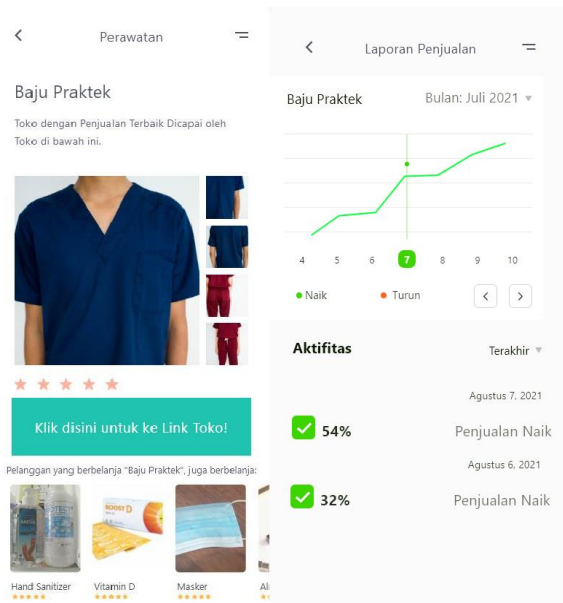


Figure 5. App overview and predictions

5. CONCLUSION

The result of this study is an application design that can help users predict and consider whether or not this product can be trusted. While speeding up the process of combining patterns of itemset combination and showing recommendations for related products to the consumers can develop the marketing of the related product, The application will improve the promotion strategy based on the product's purchase to strengthen the business. This method can help in the initial steps of improving the promotion strategy based on the product's purchase by analysing and predicting consumer behaviour in the pandemic. Based on this research, the researchers continue developing the application to make it more detailed and facilitate users with more options to enhance customer loyalty.

REFERENCES

[1] Shang, Y., Li, H., & Zhang, R. (2021). *Effect of Pandemic Outbreak on Economies:*

Evidence From Business History Context. Frontiers in Public Health, 9.

[2] Verschuur, J., Koks, E. E., & Hall, J. W. (2021). *Global economic impacts of COVID-19 lockdown measures stand out in high-frequency shipping data. PLoS ONE, 16(4).*

[3] Ozili, P., & Arun, T. (2020). *Spillover of COVID-19: impact on the Global Economy. SSRN Electronic Journal.*

[4] Puccinelli, P. J., et al. (2021) *Reduced level of physical activity during COVID-19 pandemic is associated with depression and anxiety levels: an internet-based survey. BMC Public Health, 21, 1-11.*

[5] De, R., Pandey, N., & Pal, A. (2020). *Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice. International Journal of Information Management, 55, 1-5.*

[6] Bahkir, F. A., & Grandee, S. S. (2020). *Impact of the COVID-19 lockdown on digital device-related ocular health. Indian Journal of Ophthalmology, 68, 2378-2383.*

[7] Jain, E., & Yadav, A. (2017). *Marketing and Technology: Role of Technology in Modern Marketing. IOSR Journal of Business and Management, 19(6), 49-53.*

[8] Sathya, P. (2017). *A Study on Digital Marketing and its Impact. International Journal of Science and Research, 6, 866-868.*

[9] Alford, P., & Pagi, S. J. (2015). *Marketing technology for adopting by small business. The Service Industries Journal, 35, 655-669.*

[10] Arimbawa, D. A. K., Prabawa, I. N. A., & Mertasana, P. A. (2019). *Implementation of Apriori Algorithm in Determining Tourism Visit Patterns to Bali. International Journal of*

Engineering and Emerging Technology, 4(1). 10-14.

[11] Ariestya, W. W., Supriyatni, W., & Asturi, I. (2019). *Marketing Strategy for The Determination of Staple Consumer Products Using FP-Growth and Apriori Algorithm*. *Jurnal Ilmiah Ekonomi Bisnis*, 24(3), 225-235.

[12] Raorane A A et al 2012 *Association Rule – Extracting Knowledge Using Market Basket Analysis Research Journal of Recent Sciences* pp 19-27.

[13] Adewole, K. S., Akintola, A. G., & Ajiboye, A. R. (2014). *Frequent Pattern and Association Rule Mining from Inventory Database Using Apriori Algorithm*. *African Journal of Computing & ICT*, 7(3), 35-42.

[14] Rathod, A., Dhabariya, A., & Thacker, C. (2014). *A Survey on Association Rule Mining for Market Basket Analysis and Apriori Algorithm*. *International Journal of Research in Advent Technology*, 2(3), 230-234.

[15] Safitri, A., & Purnomo, H. D. (2019). *Consumer Behavior Analysis using Apriori Algorithm*. *International Journal of Information Technology and Business*, 1(2), 1-4.

[16] Alfianzah, R., Handayani, R. I., & Murniyati. (2020). *Implementation of Apriori Algorithm Data Mining for Increase Sales*. *SinkrOn: Jurnal dan Penelitian Teknik Informatika*, 5(1), 17-25.

[17] Banerjee, K. (2018). *Enhancing Customer Loyalty using Market basket analysis, Big Data and Analytics Customer experience, customer sentiment analysis, Enterprise Analytics, market basket analysis*.

[18] Musungwini, S., Zhou, T.G., Rebanowako, T.G., Gumbo, R. & Mzikamwi, T. (2014). *The Relationship between 4Ps and Market Basket Analysis: A Case Study of Grocery Retail Shops in*

Gweru Zimbabwe. International Journal of Scientific and Technology Research

[19] Annie, L. C. & Kumar, A. D. (2012). *Market basket analysis for a Supermarket based on Frequent Itemset Mining*. *International Journal of Computer Science Issues*, 9(3), 257-264.

[20] Siahaan, H., Mawngkang, H., Efendi, S., Wanto, A., & Windarto, A. P. (2019). *Application of Classification Method C4 . 5 on Selection of Exemplary Teachers*. *IOP Conference Series*. 1–6.

[21] Kaur, M., & S. Kang. (2016). *Market basket analysis: Identify the changing trends of market data using association rule mining*. *Procedia Computer Science*, 85, 78–85.

[22] IoTUK. (2017). *Internet of Things in Retail, How can SME vendors encourage retailers to adopt the latest technologies?* <https://iotuk.org.uk/wp-content/uploads/2017/03/IoT-in-Retail.pdf>.

[23] Annie, L. C., & Kumar, A. D. (2012). *Market basket analysis for a supermarket based on frequent itemset mining*. *International Journal of Computer Science Issues*, 9(3).

[24] Venkachari, K., & Chandrasekaran, I. D. (2016). *Market basket analysis using FP Growth and Apriori Algorithm : A case study of Mumbai Retail Store*. *BVIMSR's Journal of Management Research*, 8.

[25] Setiawan, A., Budhi, G. S., Setiabudi, D. H., & Djunaidy, R. (2017). *International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT) Denpasar Bali Indonesia*, 337–340.

[26] Boztug, Y., & Reutterer, T. (2008). *A combined approach for segment-specific market basket analysis*. *European Journal of Operational Research*, 187(1). 294-312.

[27] Chen, Y. L., Tang, K., Shen, R. J., & Hu, Y. H. (2005). *Market basket analysis in a multiple store environment* *Decis. Support Syst.*

[28] Han, J., Pei, J., Mortazavi-Asl, B., Chen. Q., Dayal, U., & Hsu, M. C. (2000). *FreeSpan:*

Frequent pattern-projected sequential pattern mining Proceeding of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

[29] Yanrong, L., & Raj, G. (2005). *Effective Sampling for Mining Association Rules Ai 2004*, 391–401

[30] Wu, X., Kumar, V., Quilan, J. R., Ghosh. J., Yang, Q., & Motoda, H. (2007). *Top 10 Algorithms in Data Mining. Springer-Verlay London Limited, 14, 1-37.*