

Instagram's Apriori Association Rules Approach to Library Visitors

Egi Abinowi¹, Aminudin², Irfan Hadian Fathurrahman³, Maryam Risqa Andiani⁴

^{1,2,3,4} Widyatama University, Indonesia

ABSTRACT: The development of information technology and social media has brought significant changes in various aspects of life, including in the world of education and library management. Instagram, as one of the most popular social media platforms, has great potential to be utilized in increasing library visitor numbers. This study aims to analyze the pattern of relationships or associations between promotional activities through Instagram and library visitation rates using the Association Rules approach. This method allows for the discovery of interesting patterns from historical data that were not previously identified directly. In this study, the data analyzed was from the library's Instagram account. The results of the study show a significant correlation between certain types of posts and an increase in the number of library visitors. Through the Association Rules approach, it was found that consistent customer service with varied content and services can increase the attractiveness of Instagram users to visit the library directly. These findings are expected to serve as a reference for library managers in designing more effective and targeted digital promotion strategies, thereby increasing reading interest and active community participation in utilizing library facilities.

KEYWORDS: Apriori, Association Rules, Library

I. INTRODUCTION

The rapid development of information technology has had a significant impact on various aspects of human life, including education and information management. Social media, particularly Instagram, has become one of the most popular means of communication, promotion, and information dissemination. Instagram offers a variety of engaging features for sharing information through images, videos, and direct interaction, giving libraries significant potential to attract public attention, especially the younger generation. Unfortunately, many libraries still rely on traditional promotional methods that are less effective in today's digital age, resulting in suboptimal user visits.

To address this, a modern, data-driven approach is needed to design promotional strategies. One method that can be used is Association Rules, commonly used in market analysis to identify relationships between items in a dataset. When applied to Instagram interaction data, such as the number of likes, comments, content type, and upload time, Association Rules can help libraries understand user interest patterns. This information can be used to determine the most popular content types, the most appropriate upload times, and more targeted promotional strategies to increase visitor numbers.

In addition to providing insight into the effectiveness of digital promotions, Association Rules analysis also enables library managers to conduct periodic evaluations. In this way, the results of each promotional strategy implemented can be measured and adjusted based on real-world data. This data-driven approach is crucial to ensure that promotional efforts not only follow social media trends but also have a direct impact in the form of increased online engagement and library visits. Integrating social media with library management is expected to build a library image that is modern, engaging, and relevant to the needs of the younger generation.

This research focuses on analyzing the relationship between promotional activities on Instagram and the number of library visitors using the Association Rules method. The results are expected to provide strategic recommendations for libraries to optimize promotions through social media. Furthermore, this research is also expected to contribute to the development of knowledge in the fields of data mining, digital marketing, and library management. By utilizing appropriate technology, libraries can continue to innovate to remain relevant, adapt to current developments, and become engaging information and learning centers for the wider community.

II. METODE

The Apriori algorithm is a fundamental technique in data mining, specifically designed to discover association rules in transactional datasets. It operates on the principle of identifying frequently occurring itemsets in large datasets, thus facilitating the understanding of relationships between various data items. It is characterized by the use of two key metrics: support and confidence. Support indicates how frequently an item appears in a dataset, while confidence represents the strength of the association between items within a rule. This functionality enables businesses and researchers to extract meaningful patterns for decision-making [1].

The Apriori algorithm is a fundamental technique in data mining, specifically designed to discover association rules in transactional datasets. It operates on the principle of identifying frequently occurring itemsets in large datasets, thus facilitating the understanding of relationships between various data items. It is characterized by the use of two key metrics: support and confidence. Support indicates how frequently an item appears in a dataset, while confidence represents the strength of the association between items within a rule. This functionality enables businesses and researchers to extract meaningful patterns for decision-making [1].

Originally proposed by Agrawal et al. in 1994, the Apriori algorithm provides a systematic method for mining association rules. It uses a breadth-first search strategy to iteratively generate candidate itemsets. This approach significantly reduces computational complexity compared to naive exhaustive search, making it efficient for large database applications [2].

$$\text{support}(X \rightarrow Y) = \frac{\text{count}(X \cup Y)}{N} \quad (1)$$

Count ($X \cup Y$) = Number of transactions containing X and Y.

N = Total transactions.

However, the traditional Apriori algorithm is known for its computational intensity, particularly in the candidate generation phase, which often requires multiple passes over the dataset [3]. Numerous studies have illustrated the broad application of the Apriori algorithm across various domains. In healthcare, for example, it has been used to distinguish comorbidity patterns among patients, highlighting its effectiveness in medical data analysis [4]. Similarly, market basket analysis in retail environments has leveraged the Apriori algorithm to understand consumer purchasing behavior. By analyzing transaction data, retailers can uncover relationships between products, aiding targeted marketing and inventory management [5]. Furthermore, advances in this algorithm have resulted in modifications that improve its efficiency, such as better data structure integration that reduces the number of database scans [6].

Enhanced variants of the Apriori algorithm have addressed many of its inherent limitations. Innovations such as the Apriori_TID method optimize frequent itemset generation by mapping transactions more efficiently, significantly speeding up the mining process [7]. Furthermore, hybrid approaches incorporating machine learning techniques have been explored, further expanding the Apriori algorithm's ability to detect complex patterns in data [8].

Furthermore, the application of the Apriori algorithm continues to evolve alongside the development of data science. For example, its integration with machine learning frameworks has driven its use in predictive analytics across various sectors, including finance and environmental science. This development signals a growing trend of combining traditional data mining algorithms, such as Apriori, with modern computational techniques to handle increasingly complex data sets [9].

In short, the Apriori algorithm remains the cornerstone of association rule mining, proving vital for a variety of domains. Its structured approach to frequent itemset generation and association rule discovery not only helps businesses but also contributes significantly to research in various fields, as evidenced by its diverse applications and continued advancements [10].

The Apriori algorithm is a fundamental method in data mining, primarily used to extract frequent itemsets and generate association rules from large transactional databases. This algorithm works by leveraging the properties of support and confidence, which quantify the existence and strength of relationships between items in a dataset. Its significance is evident in various domains, particularly in shopping basket analysis, which is used to identify product purchasing patterns among consumers [5].

Developed by Agrawal et al. in 1994, the Apriori algorithm adopts a generative approach to discovering frequent itemsets. It operates based on the principle of "downward closure," which states that if an itemset is frequent, all its subsets must also be frequent. This characteristic allows the algorithm to effectively narrow the search space, significantly increasing efficiency during the mining process. However, this algorithm can be computationally intensive due to the need for multiple traversals through the dataset to identify frequent itemsets [11]. As dataset sizes increase, this inefficiency becomes more apparent, prompting researchers to refine and modify its performance [12].

The practical applications of the Apriori algorithm are numerous and diverse. In retail, these algorithms play a vital role in analyzing consumer purchasing behavior, providing actionable insights to improve sales strategies and inventory management.

III. RESULT

Each question was designed to elicit respondents' views and experiences regarding how frequently they use Instagram and the extent to which library information is conveyed through this medium. The questionnaire used a Likert scale with five answer options to facilitate measurement and analysis of the relationships between variables in this study.

The research instrument consisted of 23 questions covering: (Q1) interesting library content; (Q2) ease of information search; (Q3) freedom to choose learning resources; (Q4) communication with librarians; (Q5) quality of librarian assistance; (Q6) ease of interaction with librarians; (Q7) satisfaction with interactions with librarians; (Q8) discussions with other users; (Q9) usefulness of input from other users; (Q10) active interaction between users; (Q11) ease of receiving recommendations from other users; (Q12) productivity of information searches; (Q13) improvement in information quality; (Q14) efficiency of information searches; (Q15) flexibility of access to information sources; (Q16) productivity due to librarian interaction; (Q17) quality of work results from librarian assistance. (Q18) efficiency of consultations with librarians; (Q19) satisfaction with librarian services; (Q20) productivity due to discussions with other users; (Q21) quality of references from other users; (Q22) speed of finding references through discussions; (Q23) satisfaction after interacting with other users.

Figure 1 shows the results of data processing using the Weka application. The following parameters were used:

Minimum support: 0.3 (30 instances); Minimum metric <confidence>: 0.9; Number of cycles performed: 14; Generated sets of large itemsets: L(1)=42, L(2)=303, L(3)=698, L(4)=90.

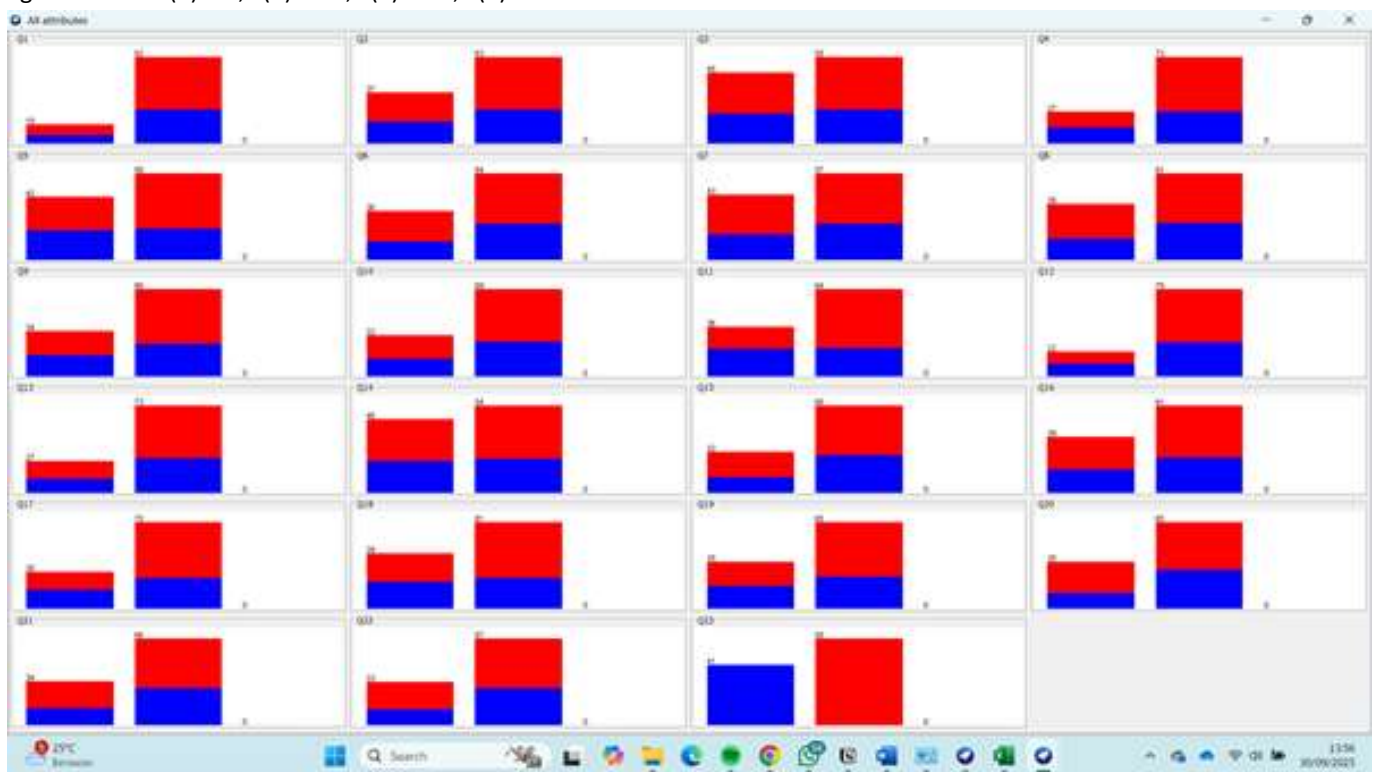


Figure 1. Results of Data Processing Using the Weka Application

Best rules found:

- (1) Q10=Netral Q19=Netral Q22=Netral 35 => Q12=Netral 34 <conf:(0.97)> lift:(1.25) lev:(0.07) [6] conv:(3.85);
- (2) Q15=Netral Q19=Netral Q22=Netral 32 => Q12=Netral 31 <conf:(0.97)> lift:(1.24) lev:(0.06) [6] conv:(3.52);
- (3) Q4=Netral Q19=Netral Q22=Netral 36 => Q12=Netral 34 <conf:(0.94)> lift:(1.21);
- (4) Q1=Netral Q19=Netral Q23=Netral 34 => Q12=Netral 32 <conf:(0.94)>;
- (5) Q4=Netral Q7=Netral Q19=Netral 34 => Q12=Netral 32 <conf:(0.94)>;
- (6) Q2=Netral Q19=Netral Q22=Netral 32 => Q12=Netral 30 <conf:(0.94)>;
- (7) Q7=Netral Q19=Netral Q22=Netral 32 => Q12=Netral 30 <conf:(0.94)>;
- (8) Q19=Netral Q22=Netral 46 => Q12=Netral 43 <conf:(0.93)>; (
- 9) Q11=Netral Q22=Netral 42 => Q1=Netral 39 <conf:(0.93)>;
- (10) Q19=Netral Q23=Netral 41 => Q12=Netral 38 <conf:(0.93)>.

The results of processing the questionnaire data using the Apriori algorithm indicate a strong correlation pattern among respondents' answers, particularly in the Neutral category. The resulting rule confirms that if a librarian answers Neutral to certain questions, they are highly likely to also answer Neutral to other related questions. For example, the rule with the highest

Instagram's Apriori Association Rules Approach to Library Visitors

confidence (97%) states that if a respondent answers Neutral to questions Q10, Q19, and Q22, they will almost certainly also answer Neutral to Q12. A similar pattern is also seen in the combination of questions Q15, Q19, and Q22 with respect to Q12, and Q4, Q19, and Q22 with respect to Q12, all with confidence levels above 94%.

These findings indicate a consistency in respondents' assessment patterns, which tend toward moderate (Neutral) scores on certain aspects, such as the quality of interactions with librarians and productivity in information retrieval. A lift value of around 1.2 indicates that this correlation is stronger than chance. Thus, libraries can interpret that groups of users who do not provide extreme ratings (strongly agree/disagree) tend to exhibit uniform attitudes on several service dimensions. This information is useful for understanding the Neutral respondent group as a target for further evaluation, as they have the potential to be the group most easily directed toward increased satisfaction with appropriate service interventions.

Q10=Neutral, Q19=Neutral, Q22=Neutral $\Rightarrow$ Q12=Neutral (conf: 0.97): If a user answers Neutral to questions about user interaction (Q10), satisfaction with librarian services (Q19), and ease of obtaining recommendations (Q22), then they are almost certainly also Neutral regarding information search productivity (Q12). This means that respondents who tend to be moderate in assessing services also perceive moderate benefits from services on their productivity.

Q15=Neutral, Q19=Neutral, Q22=Neutral $\Rightarrow$ Q12=Neutral (conf: 0.97): Neutral responses regarding flexibility of information access (Q15), satisfaction with librarians (Q19), and recommendations from other users (Q22) are strongly associated with Neutral responses regarding information productivity (Q12). This suggests that partial satisfaction with flexibility of access and librarian assistance is reflected in productivity.

Q19=Neutral, Q22=Neutral $\Rightarrow$ Q12=Neutral (conf: 0.93): The two main variables are librarian satisfaction (Q19) and recommendations from other users (Q22). If both are Neutral, then information productivity (Q12) is almost always Neutral. This is the simplest pattern that remains robust. Almost all rules involve Q12 (information search productivity) as a consequent. The factors that often appear in the antecedent are Q19 (librarian satisfaction) and Q22 (recommendations from other users), so it can be concluded that if respondents are Neutral towards librarian services and user interactions, then they are also Neutral towards information search productivity.

IV. CONCLUSIONS

Based on the Apriori results, most association rules indicate that respondents who rated librarian service (Q19) and user interaction (Q22, Q23) as Neutral tended to also give Neutral answers to information search productivity (Q12). This indicates a strong correlation between the quality of librarian service, social interaction in the library, and users' perceived productivity. Improving the quality of librarian service, more lively social interactions, and easier access to information is expected to have a more positive impact and increase the productivity of their information searches.

REFERENCES

- 1) A. H. Hasugian, M. S. Hasibuan, and S. N. Sofyan, "Apriori to Analyze Sales Patterns of Building Tools and Materials," *Journal Research and Development (ITJRD)*, vol. 7, no. 2, 2023, doi: 10.25299/itjrd.2022.10544.
- 2) P. He, B. Zhang, and S. Shen, "Effects of Out-of-Hospital Continuous Nursing on Postoperative Breast Cancer Patients by Medical Big Data," *J Healthc Eng*, vol. 2022, 2022, doi: 10.1155/2022/9506915.
- 3) R. Belmahdi, D. Mechta, and S. Harous, "A survey on various methods and algorithms of scheduling in fog computing," *International Information and Engineering Technology Association*, 2021, doi: 10.18280/isi.260208.
- 4) S. Cha and S. S. Kim, "Comorbidity patterns of mood disorders in adult inpatients: Applying association rule mining," *Healthcare (Switzerland)*, vol. 9, no. 9, Sep. 2021, doi: 10.3390/healthcare9091155.
- 5) B. H. Situmorang, A. Isra, D. Paragya, and D. A. A. Adhieputra, "Apriori Algorithm Application for Consumer Purchase Patterns Analysis," *Komputasi: Jurnal Ilmiah Ilmu Komputer dan Matematika*, vol. 21, no. 1, pp. 15–20, Jan. 2024, doi: 10.33751/komputasi.v21i1.9260.
- 6) K. Shen et al., "Association rule mining of air quality through an improved Apriori algorithm: A case study in 244 Chinese cities," *Transactions in Gis*, vol. 28, no. 4, pp. 726–745, Mar. 2024, doi: 10.1111/tgis.13156.
- 7) I. M. D. P. Asana et al., "Aplikasi Data Mining Asosiasi Barang Menggunakan Algoritma Apriori-TID," *Informatics Journal*, vol. 7, no. 1, p. 38, Apr. 2022, doi: 10.19184/isj.v7i1.30901.
- 8) R. Sarkar and J. Sun, "Improved Apriori Method for Safety Signal Detection Using Post-Marketing Clinical Data," *Mathematics*, vol. 12, no. 17, p. 2705, Sep. 2024, doi: 10.3390/math12172705.

Instagram's Apriori Association Rules Approach to Library Visitors

- 9) L. Chen, D. Xie, and B. Wang, "Research on the application of data mining technology in the evaluation and decision of network security level protection," in Third International Conference on EIEDP 2024, SPIE, Jul. 2024, p. 61, doi: 10.1117/12.3032855.
- 10) S. Qu et al., "Matchmaking in reward-based crowdfunding platforms: a hybrid machine learning approach," International Journal of Production Research, vol. 60, no. 24, pp. 7551–7571, Sep. 2022, doi: 10.1080/00207543.2022.2121870.
- 11) S. Dasgupta and B. Saha, "Study of Various Parallel Implementations of Association Rule Mining Algorithm," American Journal Of Advanced Computing, vol. 1, no. 3, pp. 1–7, Jul. 2020, doi: 10.15864/ajac.1305.
- 12) Y. Zakur and L. Flaih, "Apriori Algorithm and Hybrid Apriori Algorithm in the Data Mining: A Comprehensive Review," in ICENIS 2023, E3S Web of Conferences, Nov. 2023, p. 02021, doi: 10.1051/e3sconf/202344802021.



There is an Open Access article, distributed under the term of the Creative Commons Attribution – Non Commercial 4.0 International (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.