

Application of the Apriori Algorithm for Discovering Patterns of Success Factors in Drug Rehabilitation: A Case Study at MAB Rehabilitation Center

Regita Ramadhanty^{1*}, Magdalena Simanjuntak², Suci Ramadani³

^{1,2,3} Program of Study Information System, STMIK Kaputama
regitaaramadhanty@gmail.com^{1*}, magdalena.simanjuntak84@gmail.com², suci.ramadani23@gmail.com³

Abstract

Drug abuse is a serious issue that requires effective rehabilitation programs. MAB Rehabilitation Center possesses rehabilitation data that have mainly been used for administrative purposes and have not been analyzed to identify patterns associated with rehabilitation success. This study aims to identify relationship patterns among factors influencing rehabilitation outcomes using the Apriori Algorithm, determine the main factors forming association rules based on support, confidence, and lift ratio, and provide recommendations for rehabilitation strategies. The dataset consists of client rehabilitation records from 2020 to 2025, including variables of drug type, frequency of drug use, dependency level, therapy attendance, family support, and rehabilitation success status. Data were processed using Google Colab with a minimum support threshold of 15% and a minimum confidence threshold of 60%. The analysis generated several association rules related to rehabilitation success. The strongest rule was Moderate Dependency Level → Successful Rehabilitation, with a support value of 22.38%, a confidence value of 63.70%, and a lift ratio of 1.05, indicating the strongest association among the identified patterns. These findings demonstrate that the Apriori Algorithm is effective in discovering meaningful association patterns and can support data-driven decision-making, client assessment, and the development of more targeted rehabilitation strategies at MAB Rehabilitation Center.

Keywords: *Apriori, Data Mining, Rehabilitation Success Factors.*

1. Introduction

Drug abuse remains a serious public health and social issue in Indonesia, making rehabilitation an essential approach to restoring the physical, psychological, and social well-being of drug users. Rehabilitation centers routinely collect client data, including drug type, frequency of drug use, dependency level, therapy attendance, family support, and rehabilitation outcomes. These data have the potential to provide valuable information for identifying factors associated with successful rehabilitation; however, they are generally used only for administrative purposes and have not been fully utilized for data-driven analysis [1].

Data mining has become an effective approach for extracting meaningful knowledge from large datasets by identifying hidden patterns and relationships among variables [2] [3]. Among various data mining techniques, the Apriori Algorithm is widely used in Association Rule Mining to discover frequent itemsets and generate association rules based on minimum support and confidence thresholds [4] [5]. This study applies the Apriori Algorithm to rehabilitation data collected from MAB Rehabilitation Center to identify factors associated with successful rehabilitation. The resulting association rules are expected to support rehabilitation planning and improve data-driven decision-making within the rehabilitation center.

2. Research methodology

This study adopted the Knowledge Discovery in Database (KDD) methodology as a systematic approach to identify factors associated with the successful rehabilitation of drug users using the Apriori Algorithm. The KDD process consists of five stages: data selection, preprocessing, data transformation, data mining, and evaluation [6].

The research began with collecting rehabilitation data from MAB Rehabilitation Center, followed by data processing to generate association rules that support data-driven decision-making. The overall research methodology based on the Knowledge Discovery in Databases (KDD) framework is illustrated in Fig. 1

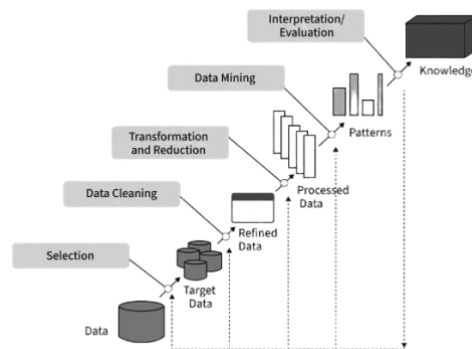


Fig. 1: Research Methodology Based on the Knowledge Discovery in Databases (KDD) Framework

2.1. Data Selection

The data selection stage involved selecting relevant rehabilitation records from MAB Rehabilitation Center covering the period 2020–2025. The dataset consisted of six variables: drug type, frequency of drug use, dependency level, therapy attendance, family support, and rehabilitation success status. These variables were selected because they are directly related to identifying factors associated with rehabilitation success.

2.2. Preprocessing

During the preprocessing stage, incomplete, duplicate, and inconsistent records were removed to improve data quality. This process ensured that the dataset was suitable for analysis and improved the reliability of the association rules generated by the Apriori Algorithm.

2.3. Data Transformation

The cleaned dataset was transformed into a transactional format suitable for Association Rule Mining. Each attribute was converted into categorical items so that it could be processed by the Apriori Algorithm to generate frequent itemsets and association rules.

2.4. Data Mining

The data mining process was performed using Google Colab by implementing the Apriori Algorithm [4] [5]. The analysis employed a minimum support threshold of 15% and a minimum confidence threshold of 60% to identify frequent itemsets and generate association rules representing relationships among factors associated with rehabilitation success.

2.5. Evaluation

The generated association rules were evaluated using support, confidence, and lift ratio [5]. The rule with the highest lift ratio was selected as the strongest association and interpreted to identify the factors most closely related to rehabilitation success. The results were then used as the basis for developing recommendations to support rehabilitation strategies and data-driven decision-making at MAB Rehabilitation Center.

3. Results and Discussion

3.1. Data Collection Results

The dataset used in this study was obtained from MAB Rehabilitation Center, consisting of rehabilitation records from 2020–2025. The collected data include six variables: drug type, frequency of drug use, dependency level, therapy attendance, family support, and rehabilitation status. A total of 800 rehabilitation records were used for the analysis after the data collection process.

Table 1: Dataset Description

Item	Description
Data Source	MAB Rehabilitation Center
Period	2020–2025
Total Records	800
Number of Variables	6

The collected dataset consists of 800 rehabilitation records with six variables used in the analysis, as presented in Table 2. These variables include Drug Type, Frequency of Drug Use, Dependency Level, Therapy Attendance, Family Support, and Rehabilitation Status. The dataset represents clients with various rehabilitation characteristics and outcomes, providing sufficient information for identifying association patterns using the Apriori Algorithm. Before the analysis, the collected data were examined to identify incomplete, duplicate, and inconsistent records before proceeding to the preprocessing stage.

Table 2: Variables Used in the Study

Variable	Description
Drug Type	Type of drug used by the client
Frequency of Drug Use	Frequency of drug use before rehabilitation
Dependency Level	Level of drug dependency
Therapy Attendance	Client attendance during rehabilitation
Family Support	Family support received by the client

3.2. Preprocessing Results

The preprocessing stage was performed to improve the quality of the rehabilitation dataset before applying the Apriori Algorithm. During this stage, the rehabilitation records were examined to identify incomplete, duplicate, and inconsistent data. Data validation was also conducted to ensure that each variable contained consistent categorical values. Furthermore, categorical variables were encoded into predefined codes to facilitate the data transformation and association rule mining processes.

3.3. Data Transformation Results

The data transformation stage was performed to convert the preprocessed rehabilitation dataset into a transactional format suitable for the Apriori Algorithm. Each rehabilitation record was represented as a combination of coded items corresponding to drug type (Z), frequency of drug use (F), dependency level (T), therapy attendance (KT), family support (DK), and rehabilitation status (S). The transformed dataset was then used as the input for frequent itemset generation and association rule mining.

Table 3: Example of Transformed Transaction Data

No	Z	F	T	KT	DK	S
1	Z2	F1	T1	KT5	DK1	S1
2	Z3	F2	T2	KT3	DK1	S0
3	Z1	F2	T3	KT1	DK2	S1
4	Z3	F3	T2	KT2	DK1	S0
5	Z1	F3	T3	KT3	DK2	S1

The transformed transaction data shown in Table 3 illustrate how each rehabilitation record was represented using predefined item codes. This transaction format served as the input for the Apriori Algorithm to generate frequent itemsets and association rules.

3.4. Association Rule Results

After the data transformation process was completed, the transaction dataset was analyzed using the Apriori Algorithm in Google Colab. The algorithm was implemented with a minimum support threshold of 15% and a minimum confidence threshold of 60% to identify meaningful association rules among the rehabilitation variables. Based on these parameters, the Apriori Algorithm generated seven association rules that satisfied the predefined thresholds.

Table 4: Association Rules Generated by the Apriori Algorithm

No	Association Rule	Support (%)	Confidence (%)	Lift
1	Moderate Dependency Level → Successful Rehabilitation	22.38	63.70	1.05
2	Ecstasy → Successful Rehabilitation	20.38	62.93	1.03
3	Adequate Family Support → Successful Rehabilitation	24.38	62.90	1.03
4	Drug Use Frequency ≥ 3 Times per Week → Successful Rehabilitation	29.50	62.60	1.03
5	Methamphetamine (Shabu) → Successful Rehabilitation	18.12	61.70	1.01
6	Drug Use > 4 Times within 2 Hours → Successful Rehabilitation	15.75	60.58	1.00
7	Severe Dependency Level → Successful Rehabilitation	25.37	60.24	0.99

Based on Table 4, the Apriori Algorithm generated seven association rules that satisfied the minimum support threshold of 15% and the minimum confidence threshold of 60%. Among these rules, the strongest association was *Moderate Dependency Level* → *Successful Rehabilitation*, with a support value of 22.38%, a confidence value of 63.70%, and a lift value of 1.05. This result indicates that participants with a moderate level of dependency had the highest association with successful rehabilitation outcomes among the identified patterns.

3.5 Discussion

The association rules generated by the Apriori Algorithm reveal meaningful relationships between rehabilitation factors and rehabilitation outcomes [4]. Among the seven generated rules, the rule *Moderate Dependency Level* → *Successful Rehabilitation* achieved the highest lift value (1.05), indicating the strongest association among the identified patterns. Although the lift value is only slightly greater than 1, it suggests that participants with a moderate dependency level were more likely to achieve successful rehabilitation outcomes than expected by chance.

In addition to dependency level, the generated association rules indicate that drug type, frequency of drug use, and family support are also associated with rehabilitation success [7]. These findings suggest that rehabilitation outcomes are influenced by multiple factors rather than a single variable. Therefore, rehabilitation planning should consider these factors collectively to improve treatment effectiveness and support decision-making.

The findings of this study demonstrate that the Apriori Algorithm is capable of identifying hidden patterns within rehabilitation data [8]. The generated association rules can assist rehabilitation institutions in evaluating participant characteristics, identifying dominant rehabilitation patterns, and supporting evidence-based rehabilitation planning. Consequently, the application of association rule mining provides valuable information for improving rehabilitation services and developing more effective intervention strategies.

4. Conclusion

This study applied the Apriori Algorithm to identify association patterns related to rehabilitation success using 800 rehabilitation records from the MAB Rehabilitation Center. By applying a minimum support threshold of 15% and a minimum confidence threshold of 60%, the algorithm successfully generated seven association rules that describe the relationships among drug type, frequency of drug use, dependency level, family support, and rehabilitation outcomes.

Among the generated rules, the strongest association was *Moderate Dependency Level* → *Successful Rehabilitation*, with a support value of 22.38%, a confidence value of 63.70%, and a lift value of 1.05. These findings demonstrate that the Apriori Algorithm can effectively discover meaningful patterns from rehabilitation data and provide valuable information to support data-driven decision-making, participant assessment, and rehabilitation planning at the MAB Rehabilitation Center.

References

- [1] Jeanne and Nur, "PROGRESIF: Jurnal Hukum XIX/No. 1/Juni 2025," *PROGRESIF: Jurnal Hukum*, Jun. 2025, [Online]. Available: <https://slemankab.bnn.go.id/pentingnya>
- [2] A. Ardhi Baskara, N. Maharani Piranti, M. Fahrury Romdendine Manajemen Teknologi Keimigrasian, P. Imigrasi Jalan Raya Gandul No, and K. Depok, "FRAMEWORK DATA MINING : SEBUAH SURVEI," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 3, 2025.
- [3] F. Harahap *et al.*, "Implementasi Algoritma Association Rule Dalam Mengidentifikasi Pola Peresepan Obat Pasien," *JURNAL SISTEM INFORMASI TGD*, vol. 4, no. 6, pp. 1452–1461, 2025, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsi>
- [4] I. Supriana and Luh, "Supriana, Implementasi Algoritma Apriori sebagai Association Rule Learning untuk Mengidentifikasi Pola Item Dataset Penjualan 69 Implementasi Algoritma Apriori sebagai Association Rule Learning untuk Mengidentifikasi Pola Item Dataset Penjualan," *Jurnal Buana Informatika*, Apr. 2025.
- [5] F. R. Jufri, S. Defit, and G. W. Nurcahyo, "Penerapan Algoritma Apriori dalam Menentukan Pola Penjualan Barang," *Jurnal KomtekInfo*, pp. 363–370, Sep. 2024, doi: 10.35134/komtekinfo.v11i4.583.
- [6] N. Purwati, Y. Pedliyansah, H. Kurniawan, S. Karnila, and R. Herwanto, "Komparasi Metode Apriori dan FP-Growth Data Mining Untuk Mengetahui Pola Penjualan," vol. 8, no. 2, 2023.
- [7] N. M. Harahap *et al.*, "Peranan Konselor Adiksi Dalam Pemulihan Pecandu Narkoba DiRehabilitasi Narkoba Sibolangit Centre," *Jurnal Media Intelektual Muslim dan Bimbingan Rohani*, vol. 11, no. 1, p. 2025, [Online]. Available: <http://journal.iaimsinjai.ac.id/index.php/mimbar>
- [8] M. Yudhistira, R. S. Rohman, E. Marsusanti, U. Bina, S. Informatika, and D. D. Disetjui, "Penerapan Association Rule Menggunakan Algoritma Apriori Untuk Meningkatkan Penjualan Di Kandang Kopi Tasikmalaya," *Journal Computer Science*, vol. 2, no. 2, 2023.