Indonesian Journal of Data and Science

Volume 7 Issue 1
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v7i1.398

Research Article

The Effect of Clinical Rule-Based Domain Filtering on the Performance of FP-Growth-Based Drug Recommendation Systems

Muhammad Zaqly Luluang [1]; Irawati [2]; Herdianti Darwis [3]
[1] Universitas Muslim Indonesia, Indonesia, Makassar, 13020220203@student.umi.ac.id
[2] Universitas Muslim Indonesia, Indonesia, Makassar, irawati@umi.ac.id
[3] Universitas Muslim Indonesia, Indonesia, Makassar, herdianti.darwis@umi.ac.id

Correspondence should be addressed to Muhammad Zaqly Luluang; 13020220203@student.umi.ac.id

Received 10 December 2025; Accepted 15 February 2026; Published 30 March 2026

Abstract:
This study analyzes the effect of domain filtering on drug recommendation systems based on association rule mining using the FP-Growth algorithm with Neural Collaborative Filtering (NCF) as a comparison. The dataset used was derived from patient medical records containing attributes such as complaints, diagnoses, and drug therapies, with a total of 1,000 patient transactions. To avoid data leakage, the dataset was randomly divided into 70% training data and 30% test data before the modeling process was carried out. Domain filtering was applied by limiting the rule structure so that complaints and diagnoses acted as antecedents and drugs as consequents. The performance of the recommendation system was evaluated using the Precision@5, Recall@5, and Normalized Discounted Cumulative Gain (NDCG@5) metrics. The results of the experiment show that the FP-Growth approach with domain filtering produces higher Precision@5 and NDCG@5 values than the non-filtering approach. The Wilcoxon Signed-Rank test shows that the difference is statistically significant, while effect size analysis using Cliff's Delta shows a practically meaningful impact. Furthermore, a comparison with Neural Collaborative Filtering shows that the collaborative filtering-based approach is less effective on transactional clinical prescription data with limited historical interactions. These findings indicate that integrating medical domain knowledge into FP-Growth can improve the clinical relevance and quality of drug recommendation rankings.

Keywords: Association rule mining, FP-Growth, domain filtering, drug recommendation system; medical data.

Dataset Link: BPJS Drug Recommendation Dataset at the Bonto Perak Community Health Center, Pangkep Regency, South Sulawesi.

1. Introduction

The use of Electronic Medical Record (EMR) data to support clinical decision-making continues to grow, particularly in the context of drug therapy recommendations based on historical data [1], [2], [3], [4]. EMR data containing information on complaints, diagnoses, and drug therapy provide a structured source of clinical data that can be used to help medical personnel determine drug recommendations in a more consistent and data-driven manner [5], [6]. However, the complexity and heterogeneity of clinical data means that the patterns generated can be difficult to interpret if not processed using an approach that is appropriate to the clinical context [7], [8], [9].

Association Rule Mining (ARM) is one of the most widely used approaches for extracting patterns of association from medical transactional data, with the FP-Growth algorithm often chosen for its efficiency in handling large-scale data [10], [11]. This approach has been applied to medical and pharmaceutical data to identify patterns of drug use and prescription. However, the application of ARM without domain restrictions on clinical data often results in a large number of rules that are statistically valid but clinically meaningless (rule explosion), thereby limiting its direct use in drug recommendation systems [12].

In developing a clinical data-based drug recommendation system, transparency and consistency with medical reasoning are important because the recommendations generated can influence clinical decisions [13], [14], [15]. Therefore, the association rules-based approach is still relevant due to its transparent and traceable nature, provided that the rules generated are aligned with medical domain knowledge [16]. Restrictions on the structure of the rules, such as defining complaints and diagnoses as antecedents and drugs as consequents, are necessary so that the drug recommendations generated are more clinically relevant and can be used effectively in practice.

Previous studies have shown that Association Rule Mining (ARM) algorithms, particularly FP-Growth, have been widely applied in the pharmaceutical domain to analyze drug purchasing patterns based on pharmacy sales transaction data. Several studies have utilized FP-Growth to identify frequent itemsets of drugs that are often purchased together to support stock management and drug layout planning [17], [18]. Meanwhile, other studies have applied FP-Growth-based association methods to large-scale transaction datasets as a basis for procurement planning and decision-making related to drug stocks [19]. Although effective in extracting statistical patterns from transaction data, most of these studies still view drugs as commercial items in a shopping basket (market basket analysis), so that the resulting association rules focus on the frequency of co-occurrence without considering the clinical context of patients or their direct integration into medical needs-based drug recommendation systems.

Unlike previous studies that focused on analyzing drug sales transactions, this study emphasizes the application of domain filtering in the FP-Growth algorithm to improve the clinical relevance of the resulting association rules. Rules are structurally constrained by setting complaints and diagnoses as antecedents and drugs as consequents, thus following the pattern of clinical reasoning in medical practice. The resulting rules are then implemented in a rule-based drug recommendation system to generate Top-K drug recommendations that are appropriate for the patient's condition. The system's performance is evaluated using Precision@K, Recall@K, and NDCG@K metrics and compared with the Neural Collaborative Filtering (NCF) method as a baseline to provide performance context for modern recommendation approaches. In addition, the evaluation is reinforced with non-parametric statistical tests to ensure that the improvement in clinical relevance achieved maintains the quality of recommendations significantly.

Based on these issues, this study hypothesizes that the application of domain filtering in the FP-Growth algorithm can improve the quality of drug recommendation systems, particularly in terms of recommendation accuracy and ranking quality, without reducing the clinical relevance of the resulting association rules.

2. Method

This study applies an experimental approach to build and evaluate a drug recommendation system based on association rules. The methodological flow includes processing medical transaction data, forming association rules using the FP-Growth algorithm with two scenarios (non-filtering and filtering domain), implementing rule-based drug recommendations, and evaluating performance using Top-K metrics and statistical analysis. Additionally, Neural Collaborative Filtering (NCF) is used as a comparative baseline to provide performance context for modern recommendation approaches. All stages are designed to ensure the validity of the results and avoid data leakage during the modeling and evaluation processes. The research flow can be seen in **Figure 1.**

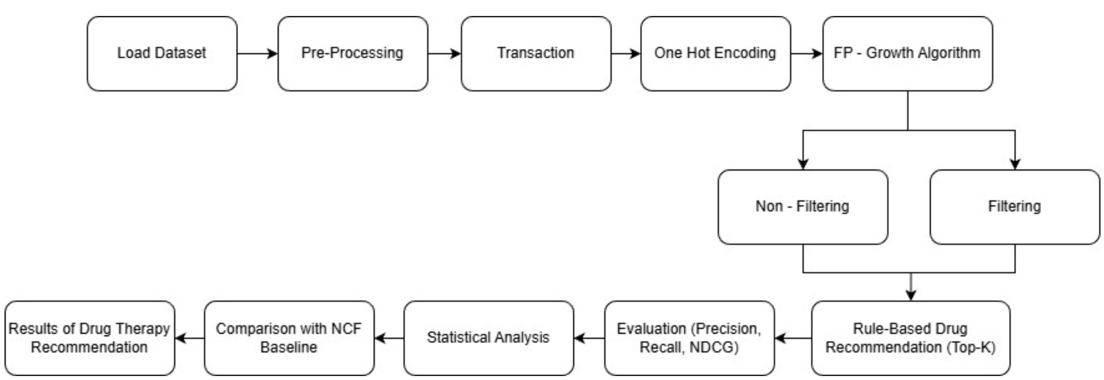

Figure 1. Research flowchart

a. Dataset and Preprocessing

The data used in this study was sourced from the medical records of BPJS participants at the Bonto Perak Community Health Center, Pangkep Regency, South Sulawesi. Of all the attributes available in the medical record data, this study only utilized the attributes of complaints, diagnoses, and drug therapy, as these three attributes are directly relevant to the objectives of association rule analysis and the development of a drug recommendation system. Other attributes not directly related to the formation of transaction items were not included in the analysis process..

Raw medical record data first undergoes preprocessing, which includes data cleaning, label normalization, and filtering of incomplete or invalid data. Next, the data is represented in the form of transactions, where each patient is treated as a single transaction that can contain more than one complaint, diagnosis, or drug therapy item. Transactions that do not meet the format or criteria for item formation are excluded from the analysis process. The research sample consists of all transactions that meet these criteria and is used as the basis for model formation and recommendation system evaluation.

To avoid data leakage, the transaction dataset is divided into a training set and a test set before the modeling process is carried out [20]. The data is divided randomly with a ratio of 70% training data and 30% test data [21]. The training data is used for the association rule formation process using the FP-Growth algorithm, while the test data is used exclusively in the recommendation system performance evaluation stage.

Table 1 shows examples of patient medical record data used in this study. Each row in the table represents one patient, while the main columns consist of complaints, diagnoses, and drug therapies. The complaint attribute contains the symptoms experienced by the patient, the diagnosis attribute contains the results of the diagnosis made by medical personnel, and the drug therapy attribute records the drugs prescribed to the patient during that visit.

Table 1. BPJS dataset for the Bonto Perak Community Health Center, Pangkep Regency, South Sulawesi

No	Keluhan	Diagnosa	Terapi Obat
1	['demam', 'pilek', 'sakit_kepala']	['hyperlipidaemia']	['vitamin b kompleks', 'simvastatin']
2	['batuk', 'demam', 'pilek']	['fever']	['vitamin c', 'sefadroksil', 'klorfeniramin', 'parasetamol']
…	…	…	…
999	['sakit_kepala']	['essential (primary) hypertension', 'headache']	['amlodipin', 'parasetamol', 'sianokobalamin vitamin b12']
1000	['leher_tegang']	['non-insulin-dependent diabetes mellitus without complications', 'hyperuricaemia without signs of inflammatory arthritis and tophaceous disease', 'pain in joint']	['omeprazole', 'alopurinol', 'ibuprofen', 'vitamin b kompleks']

b. Transaction Encoding and Association Rule Mining

Transaction data in the training data is then represented in binary format using the one-hot encoding technique. This representation allows each item in the transaction to be expressed as a binary value, so that it can be processed by the association rule mining algorithm [22]. The encoding process is applied consistently to the training data, while the test data is stored separately for the evaluation stage [23].

The FP-Growth algorithm is used to extract frequent itemsets from transaction data without the need to explicitly generate itemset candidates [24]. Based on the frequent itemsets obtained, association rules are formed by considering statistical measures commonly used in ARM [25]. In this study, two FP-Growth scenarios were applied. The first scenario is non-filtering FP-Growth, where all item combinations in transactions are considered without rule structure restrictions. The second scenario is FP-Growth with domain filtering, where rules are structurally constrained based on clinical logic, i.e., only complaints and diagnoses are allowed as antecedents and drugs as consequents. This approach aims to reduce the number of clinically irrelevant rules and improve the interpretability of the generated rules.

Table 2. Results of One-Hot Encoding of Medical Transaction Data

No	d_"abdominal _pain"	d_"abscess"	d_"acute _abdomen"	...	o_'vitamin _b_kompleks'	o_ 'vitamin_c'	o_ 'zinc'
1	0	0	0	...	0	0	0
2	0	0	0	...	0	1	0
3	0	0	0	...	0	0	0
4	0	0	0	...	0	0	0
5	0	0	0	...	0	1	0
…	…	…	…	…	…	…	…
698	0	0	0	...	0	0	0
699	0	0	0	...	0	0	0
700	0	0	0	...	0	0	0
701	0	0	0	...	0	0	0
702	0	0	0	...	0	0	0

Table 2 shows the results of transforming medical transaction data using the one-hot encoding technique, in which categorical data is converted into binary representation. The rows in the table represent patient transactions, while the columns show the normalized complaint, diagnosis, and medication items. This encoding process is applied to the training data to support the formation of frequent itemsets in the FP-Growth algorithm.

c. Rule- Based Drug Recommendation

The association rules generated from FP-Growth are used as the basis for a rule-based drug recommendation system [26]. For each transaction in the test data, the system matches the patient's condition with the available antecedent rules. Drugs that appear in the corresponding consequent rules are collected as recommendation candidates. Next, these drug candidates are ranked based on the strength of the underlying rules.

The recommendation system generates a list of drug recommendations in the form of Top-K, where the value of K indicates the maximum number of drugs recommended for each patient case. This approach is applied consistently in both FP-Growth scenarios, namely non-filtering and filtering domains, thereby enabling a fair comparison of performance between the two approaches in generating drug recommendations.

In the non-filtering approach, association rules can generate consequences that are not always drugs because all items in the transaction are considered without domain restrictions. This condition reflects the basic characteristics of association rule mining, which extracts all patterns of association in the data. Conversely, in the domain filtering approach, the consequences are limited to only drug items so that the rule structure is more in line with clinical logic and more suitable for use in drug recommendation systems.

d. Evaluation and Comparison with Baseline

The performance of the recommendation system was evaluated using the Top-K approach commonly used in recommendation system research [27]. The evaluation was performed on test data to assess the quality of drug recommendations generated by each approach. The evaluation metrics used include Precision@K, Recall@K, and Normalized Discounted Cumulative Gain (NDCG@K), which are widely used to evaluate the relevance and ranking of recommendations [28].

To analyze the performance differences between the non-filtering and filtering domain FP-Growth approaches, the Wilcoxon Signed-Rank Test non-parametric statistical test was used. In addition, the effect size was calculated using Cliff's Delta to measure the magnitude of the influence of domain filtering application. As a comparison, Neural Collaborative Filtering (NCF) was used as a modern recommendation system baseline, which is part of a recommendation system approach that models interactions between users and items to generate predictions based on historical data [29]. The recommendation results generated by NCF were evaluated using the same metrics and compared with the FP-Growth-based approach to provide performance context for neural network-based recommendation methods.

In the implementation of Neural Collaborative Filtering, patients are represented as users and drugs as items in the user-item interaction matrix. Interactions are recorded when a drug is prescribed to a patient in historical data. The NCF model is trained using an embedding layer followed by a multilayer perceptron to model the non-linear relationship between users and items [30]. The training process uses a binary cross-entropy loss function with an Adam optimizer and negative sampling techniques to form negative interaction pairs from drugs that are not prescribed to patients.

e. Evaluation Metrics and Statistical Formulation

The evaluation of the drug recommendation system's performance in this study focused on the system's ability to generate a list of drugs relevant to the patient's clinical condition. Since the recommendation system generates a set of items in the form of a ranked list, the evaluation approach used must be able to assess the accuracy, completeness, and quality of the recommendation rankings [31]. Therefore, this study uses Top-K-based evaluation metrics commonly used in recommendation system research, namely Precision@K, Recall@K, and Normalized Discounted Cumulative Gain (NDCG@K).

Precision@K is used to measure the accuracy of recommendations by calculating the proportion of relevant drugs among the top K recommended drugs [32]. In a clinical context, this metric is important because it reflects the extent to which the system is able to minimize irrelevant drug recommendations, thereby potentially reducing the risk of medication errors.

Recall@K is used to measure the system's ability to cover all relevant drugs that should be recommended to patients [33]. This metric is important to ensure that the system does not ignore relevant therapies based on historical data. By considering Precision@K and Recall@K together, this study can analyze the trade-off between the accuracy and completeness of drug recommendations.

In addition, NDCG@K is used to evaluate the quality of recommendation rankings [34]. This metric considers the position of relevant items in the recommendation list, where drugs that appear higher up in the list have a greater contribution. The use of NDCG@K is relevant in this study because drug recommendation systems are not only required to produce the correct drugs, but also to place the most relevant drugs at the top of the list so that they can be easily interpreted by medical personnel.

To ensure that the difference in performance between the non-filtering and filtering domain FP-Growth approaches did not occur by chance, the Wilcoxon Signed-Rank Test [35]. non-parametric statistical test was used. This test was chosen because it does not assume a normal distribution and is suitable for comparing two approaches evaluated on the same data pairs. In addition to statistical significance, Cliff's Delta effect size was used to measure the magnitude of the effect of domain filtering on the practical quality of drug recommendations [36]. Cliff's Delta was calculated based on a comparison of paired values from the two methods being compared.

In addition to the Top-K evaluation metric, this study also uses the Average Clinical Relevance Score (ACRS) to assess the clinical relevance of drug recommendations at the individual case level. ACRS measures the strength of the relationship between a patient's clinical condition and the recommended drug by aggregating the association rules that support the recommendation.

$$Support(X) = \frac{|\{t \in T \mid X \subseteq t\}|}{|T|} \tag{1}$$

$$Confidence(X \longrightarrow Y) = \frac{support(X \cup Y)}{support(X)} \tag{2}$$

$$Precision@K = \frac{|R\kappa \cap T|}{K} \tag{3}$$

$$Recall@K = \frac{|R\kappa \cap T|}{T} \tag{4}$$

$$DCG@K = \sum_{i=1}^{K} \frac{2^{rel_i} - 1}{Log_2(i+1)} \tag{5}$$

$$NDCG@K = \frac{DCG@K}{IDCG@K} \tag{6}$$

$$W = \min(W^+, W^-) \tag{7}$$

$$\delta = \frac{\sum\sum[x_i > y_j] - [x_i < y_j]}{mn} \tag{8}$$

$$ACRS = \frac{\sum_{r\in R}\big(\alpha \cdot Confidence(r) + \beta \cdot Support(r)\big) \times Penalty(r)}{\sum_{r\in R} Support(r)} \tag{9}$$

3. Result and Discussion

This section presents the results of evaluating the performance of a drug recommendation system based on association rule mining using the proposed FP-Growth algorithm. The evaluation was conducted to compare non-filtering and filtering approaches in the medical domain, as well as to place its performance in the context of machine learning-based recommendation approaches through comparison with Neural Collaborative Filtering (NCF). Testing was performed on test data separate from the training data to avoid data leakage. System performance was evaluated using the Precision@5, Recall@5, and Normalized Discounted Cumulative Gain (NDCG@5) metrics, which represent recommendation accuracy, recommendation coverage, and the quality of recommended drug rankings, respectively.

a. Performance Results of the FP-Growth-Based Recommendation System

Evaluation The performance of the recommendation system was evaluated to compare the FP-Growth approach without filtering and FP-Growth with domain filtering at various minimum support values. The testing was conducted using test data that was separate from the training data to avoid data leakage. The evaluation was conducted using the Precision@5, Recall@5, and NDCG@5 metrics, which represent the accuracy of recommendations, the coverage of recommendations, and the quality of the recommended item rankings, respectively. The results of the FP-Growth-based recommendation system performance evaluation are presented in Table 3.

Table 3. Results of the Evaluation of the FP-Growth-Based Recommendation System

min_ support	Precision@5 _raw	Recall@5 _raw	NDCG@5 _raw	Precision@5 _domain	Recall@5 _domain	NDCG@5 _domain
0.01	0.380333333	0.361444444	0.403612543	0.603833333	0.282	0.617513922
0.02	0.359944444	0.293944444	0.366544901	0.546111111	0.218222222	0.552988803
0.05	0.362888889	0.239777778	0.364099127	0.498888889	0.173944444	0.502784633
0.1	0.344055556	0.197722222	0.314920154	0.437222222	0.149944444	0.441699933
0.15	0.210555556	0.120555556	0.185150594	0.355	0.120555556	0.356885787
0.2	0.138333333	0.071666667	0.16963739	0.276666667	0.071666667	0.276666667

The domain filtering approach consistently produces higher Precision@5 and NDCG@5 values compared to the non-filtering approach across all minimum support values tested. This shows that integrating medical domain knowledge can improve the accuracy of drug recommendations and the quality of the resulting recommendation rankings. Conversely, the non-filtering approach tends to produce slightly higher Recall@5 values at some minimum support configurations, indicating that this approach produces a broader range of recommendations, but potentially reduces clinical relevance because it does not explicitly consider medical domain constraints..

To further illustrate the effect of minimum support on recommendation quality, the relationship between minimum support and NDCG@5 is visualized in **Figure 2**. The visualization shows that the domain filtering approach consistently maintains higher and more stable NDCG@5 values compared to the non-filtering approach across the entire range of minimum support, indicating that the benefits of domain filtering are not limited to specific configurations but remain consistent across various parameter settings.

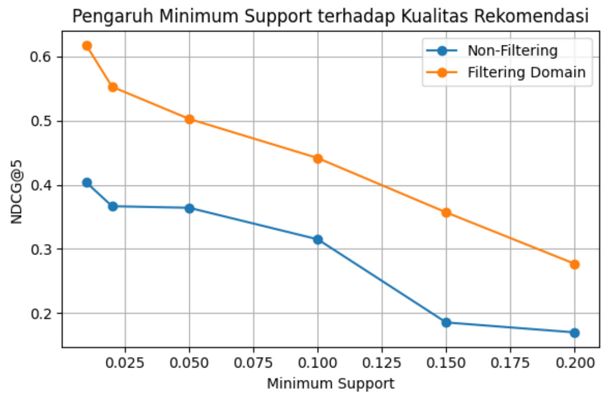

Figure 2. The Effect of Minimum Support on NDCG@5 Values

b. Statistical Significance of Recommendation System Performance

Although the performance differences in **Table 3** are evident numerically, they need to be further tested using a statistical approach to ensure that the performance improvements obtained truly reflect real differences and are not merely variations that occur by chance due to data characteristics. Therefore, a Wilcoxon Signed-Rank test was performed on the Precision@5, Recall@5, and NDCG@5 values calculated for each transaction in the same test data, so that the comparison was made in pairs (paired comparison) and was able to evaluate the performance differences between the two approaches more accurately under identical test conditions.

Table 4. Wilcoxon Test Results for Recommendation System Performance

Metric	p_value	Decision
Precision@5	2.61E-21	Significantly different
Recall@5	9.05E-11	Significantly different
NDCG@5	9.39E-20	Significantly different

The results in **Table 4** show that the difference in performance between the non-filtering and domain filtering approaches is statistically significant across all evaluation metrics ($p < 0.05$), namely Precision@5, Recall@5, and NDCG@5. This indicates that the application of domain filtering provides consistent performance improvements across various test transactions and is not caused by random variations in the data. This improvement indicates that the domain filtering mechanism is capable of improving the relevance of recommended items and enhancing the quality of drug recommendation rankings generated by the system.

Although all metrics show statistically significant differences, the effect size analysis shows that the impact of domain filtering on Recall@5 is relatively smaller than its impact on Precision@5 and NDCG@5. This indicates that the main benefit of domain filtering lies in improving the accuracy of recommendations and the quality of recommendation rankings, rather than expanding the coverage of relevant items recommended.

The visualization of the Wilcoxon test results is shown in **Figure 3**, which shows a paired comparison of NDCG@5 values before and after applying domain filtering. This visualization provides an overview of the distribution of performance value changes and reinforces the evidence that the domain filtering approach can consistently improve the quality of recommendation rankings in most test transactions.

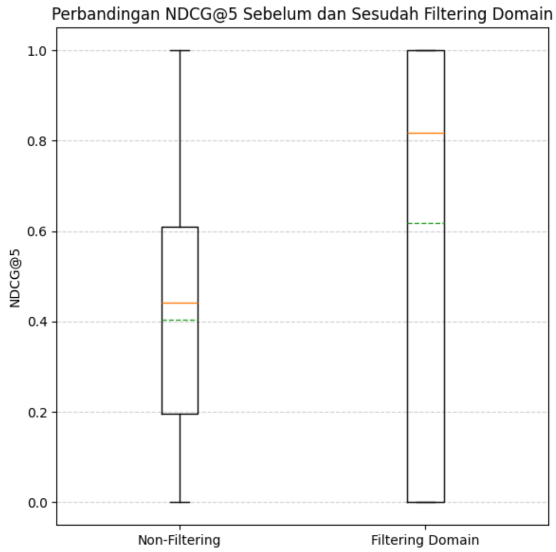

Figure 3. Boxplot Comparison of NDCG@5 Before and After Domain Filtering

c. Effect Size Analysis

In addition to testing statistical significance, an effect size analysis was conducted using Cliff's Delta to assess the magnitude of the impact of domain filtering on the practical performance of the recommendation system. This analysis complements the significance test results because the p-value only indicates a difference, while Cliff's Delta shows the strength of the effect of that difference.

In this study, Cliff's Delta values were calculated using the comparison convention (non-filtering − filtering). Therefore, negative Cliff's Delta values indicate that the FP-Growth approach with domain filtering produces better performance than the non-filtering approach.

Table 5. The Effect of Filtering Domain Size on System Performance

Metrics	Cliffs_Delta	Effect
Precision@5	-0.326	Small
Recall@5	0.198	Small
NDCG@5	-0.338	Medium

The results in **Table 5** show that the application of domain filtering has varying degrees of impact on each evaluation metric. The Cliff's Delta value in Precision@5 shows a small effect, indicating a consistent improvement in recommendation accuracy. The NDCG@5 metric shows a moderate effect size, indicating that domain filtering has a stronger influence in improving the quality of recommendation rankings. Meanwhile, the Cliff's Delta value in Recall@5 shows a small effect, indicating that changes in the coverage of relevant items recommended are relatively smaller than the improvement in the accuracy and quality of recommendation rankings.

These findings reinforce the results of the previous Wilcoxon test, which showed that the performance improvement achieved through the application of domain filtering is not only statistically significant, but also has a meaningful practical impact, particularly in improving the accuracy and quality of recommendation rankings, which are important aspects in clinical recommendation systems to ensure that the recommendations provided are more relevant, targeted, and can support the decision-making process more effectively.

d. Comparison with Neural Collaborative Filtering (NCF)

To provide context for the deep learning-based recommendation approach, a comparison was made between FP-Growth without filtering, FP-Growth with domain filtering, and Neural Collaborative Filtering (NCF). This comparison was performed using the best configuration based on the NDCG@5 value, and the results are presented in **Table 6**. As shown in the table, the FP-Growth approach with domain filtering achieved the best performance on the Precision@5 and NDCG@5 metrics, indicating higher recommendation accuracy and ranking quality compared to other approaches. On the other hand, Neural Collaborative Filtering produced lower values on all evaluation metrics, indicating that collaborative filtering is less effective for transactional clinical prescription data with limited historical interactions.

Table 6. Comparison of Best Performance Between Recommendation Methods

	Precision@5	**Recall@5**	**NDCG@5**
Non-Filtering	0.380333333	0.361444444	0.403612543
Filtering	0.603833333	0.282	0.617513922
NCF	0.160144928	0.20821256	0.2298014

The low performance of NCF indicates that the collaborative filtering approach is less suitable for transactional clinical prescription data, which has infrequent interaction patterns and is prone to cold-start problems. Limited historical interactions make it difficult for the model to learn stable relationships between patients and drugs, resulting in recommendations that are less aligned with the clinical context. In contrast, the rule-based approach with domain filtering is able to utilize the direct relationship between complaints, diagnoses, and drugs, resulting in recommendations that are more relevant and consistent with the patient's condition.

The visual comparison in **Figure 4** focuses on the Precision@5 and NDCG@5 metrics because these two metrics represent the accuracy and quality of recommendation rankings. Although Recall@5 was also evaluated, this metric is not shown in the main visualization because domain filtering produces more selective recommendations, so improvements are more noticeable in accuracy and ranking quality than in recommendation coverage. This indicates that domain filtering is more effective in improving the relevance of recommendations in a clinical context.

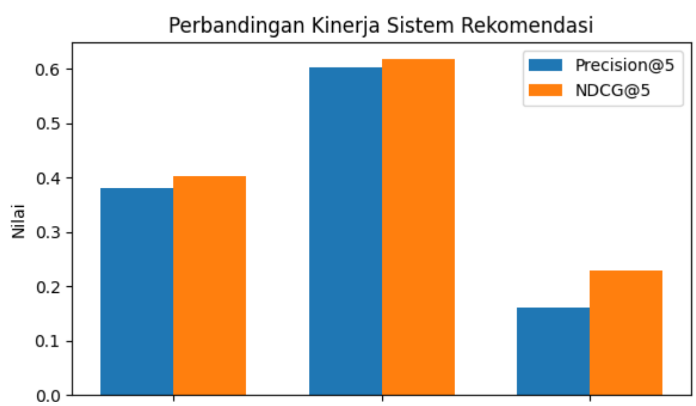

Figure 4. Comparison of Recommendation System Performance on Precision@5 and NDCG@5

e. Case Study Analysis and Clinical Relevance

In addition to quantitative evaluation, a case study analysis was conducted to assess the clinical relevance of recommendations at the individual level using the Average Clinical Relevance Score (ACRS). The ACRS value is calculated based on the association rules that support drug recommendations in a patient case. This score reflects the strength of the relationship between the patient's clinical condition and the recommended drug, taking into account the confidence and support values, as well as the suitability of the rules with the medical domain structure. A higher ACRS value indicates that the drug recommendation has a stronger relationship with the patient's clinical condition based on patterns found in historical data. The ACRS value can vary depending on the number and strength of the association rules that support a drug recommendation. The more relevant rules there are and the higher the confidence and support values of those rules, the higher the resulting ACRS value will be. Examples of recommendation results and ACRS values are presented in **Table 7**.

Table 7. Case Study Example: Drug Recommendations and ACRS Values

Case ID	Complaints	Diagnosis	Recommendation Type	Recommended Drugs	Clinical Relevance Score
139	Joint pain	low back pain	Filtering Domain	sianokobalamin vitamin b12	42.4
139	Joint pain	low back pain	Non-Filtering	nyeri sendi, sianokobalamin vitamin b12	7.798
215	Cough	acute nasopharyngitis, essential (primary) hypertension	Filtering Domain	amlodipin, asetil sistein, klorfeniramin, parasetamol	6.104
37	Cough, fever, cold	acute upper respiratory infection	Filtering Domain	amoxicilin, klorfeniramin, parasetamol	5.081
215	Cough	acute nasopharyngitis, essential (primary) hypertension	Non-Filtering	essential (primary) hypertension, batuk, amlodipin, asetil sistein, klorfeniramin	4.186

The difference in ACRS scores between the domain filtering and non-filtering approaches shows that restricting the structure of the rules can strengthen the relationship between the patient's clinical condition and the recommended medication. The domain filtering approach produces higher ACRS values, indicating that the resulting drug recommendations are more clinically relevant. In addition, this approach avoids the appearance of non-drug items in the recommendation list, making the resulting recommendations easier for medical personnel to interpret.

f. Implications of Domain Filtering Application in Clinical Decision Support System Development

Overall, the results of the study show that the integration of medical domain knowledge through the domain filtering mechanism in the FP-Growth algorithm can significantly improve the quality of the drug recommendation system, particularly in terms of the accuracy and quality of recommendation rankings. By limiting the structure of the rules to align with clinical logic, the system can generate more relevant recommendations and reduce the occurrence of medically meaningless rules. Furthermore, this approach maintains the interpretability of association rules, allowing healthcare professionals to transparently trace and understand the relationship between a patient's clinical condition and drug recommendations.

The association rule-based approach enriched with domain filtering is considered more suitable for clinical settings, especially in conditions with limited historical data and the need for transparency in decision making. Unlike the collaborative filtering approach, which relies on historical interaction patterns, this approach utilizes the direct relationship between complaints, diagnoses, and drug therapies, making it more stable and relevant in a clinical context. Therefore, the application of domain filtering has the potential to be integrated into clinical decision support systems to assist medical personnel in obtaining more accurate, relevant, and clinically accountable drug recommendations.

4. Conclusion

This study analyzes the effect of domain filtering on drug recommendation systems based on association rule mining using the FP-Growth algorithm with Neural Collaborative Filtering as a comparison. Domain filtering is applied by limiting the rule structure so that complaints and diagnoses act as antecedents and drugs as consequents. Evaluation uses Precision@5, Recall@5, and NDCG@5 metrics on separate test data to avoid data leakage.

The results of the experiment show that the FP-Growth approach with domain filtering consistently improves Precision@5 and NDCG@5 compared to the non-filtering approach. The Wilcoxon Signed-Rank test shows that the difference is statistically significant, while Cliff's Delta analysis shows the practical impact of applying domain filtering. A comparison with Neural Collaborative Filtering shows that the collaborative filtering-based approach is less effective on transactional clinical prescription data and has limitations in terms of historical interactions.

This study shows that the application of domain filtering on FP-Growth can improve the clinical relevance and quality of drug recommendations. However, this study has limitations because the dataset used came from a single healthcare facility. Further research could conduct cross-facility evaluations and involve direct validation by medical personnel to improve the generalization and reliability of the proposed system.

References:

[1] S. Murthi, N. Martini, N. Falconer, and S. Scahill, “Evaluating EHR-Integrated Digital Technologies for Medication-Related Outcomes and Health Equity in Hospitalised Adults: A Scoping Review,” Dec. 01, 2024, *Springer*. doi: 10.1007/s10916-024-02097-5.

[2] C. Scherkl *et al.*, “Towards a prescribing monitoring system for medication safety evaluation within electronic health records: a scoping review,” *BMC Med. Inform. Decis. Mak.*, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12911-025-03096-3.

[3] L. Syafie and N. Rismayanti, “Automated Diagnosis of Benign Prostatic Hyperplasia Using Deep Learning on RGB Prostate Images,” *International Journal of Artificial Intelligence in Medical Issues*, vol. 3, no. 1, pp. 41–50, May 2025, doi: 10.56705/ijaimi.v3i1.251.

[4] M. F. Banjar, I. Irawati, F. Umar, and L. N. Hayati, “Analysis of Stroke Classification Using Random Forest Method,” *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 186–193, Dec. 2022, doi: 10.33096/ilkom.v14i3.1252.186-193.

[5] S. E. Davis *et al.*, “Use of Electronic Health Record Data for Drug Safety Signal Identification: A Scoping Review,” Aug. 01, 2023, *Adis*. doi: 10.1007/s40264-023-01325-0.

[6] I. As’ad, “Advancing Healthcare Diagnostics: A Study on Gaussian Naive Bayes Classification of Blood Samples,” *International Journal of Artificial Intelligence in Medical Issues*, vol. 1, no. 2, pp. 115–123, Nov. 2023, doi: 10.56705/ijaimi.v1i2.120.

[7] A. Q. Wang *et al.*, “A Framework for Interpretability in Machine Learning for Medical Imaging,” *IEEE Access*, vol. 12, pp. 53277–53292, 2024, doi: 10.1109/ACCESS.2024.3387702.

[8] Y. Salim, A. P. Utami, A. R. Manga, H. Azis, and F. T. Admojo, “Optimal Strategy for Handling Unbalanced Medical Datasets: Performance Evaluation of K-NN Algorithm Using Sampling Techniques,” *Knowledge Engineering and Data Science*, vol. 7, no. 2, Dec. 2024, doi: 10.17977/um018v7i22024p176-186.

[9] A. Rachman Manga, A. Putri Utami, H. Azis, Y. Salim, and A. Faradibah, “Optimizing classification models for medical image diagnosis: a comparative analysis on multi-class datasets,” *Computer Science and Information Technologies*, vol. 5, no. 3, pp. 205–214, 2024, doi: 10.11591/csit.v5i3.pp205-214.

[10] H. Yamamoto, G. Kayanuma, T. Nagashima, C. Toda, K. Nagayasu, and S. Kaneko, “Early Detection of Adverse Drug Reaction Signals by Association Rule Mining Using Large-Scale Administrative Claims Data,” *Drug Saf.*, vol. 46, no. 4, pp. 371–389, Apr. 2023, doi: 10.1007/s40264-023-01278-4.

[11] H. Harlinda and R. Satra, "Data Mining Approach to Improve Minimarket Sales using Association Rule Method," *Jurnal Informatika*, vol. 12, no. 1, pp. 10–14, 2025, doi: 10.31294/inf.v12i1.20835.

[12] S. Darrab, D. Broneske, and G. Saake, "Exploring the predictive factors of heart disease using rare association rule mining," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-69071-6.

[13] S. Hur *et al.*, "Comparison of SHAP and clinician friendly explanations reveals effects on clinical decision behaviour," *NPJ Digit. Med.*, vol. 8, no. 1, Dec. 2025, doi: 10.1038/s41746-025-01958-8.

[14] M. R. Pinsky *et al.*, "Use of artificial intelligence in critical care: opportunities and obstacles," *Crit. Care*, vol. 28, no. 1, Dec. 2024, doi: 10.1186/s13054-024-04860-z.

[15] P. Lestari, L. Belluano, R. A. Rahma, H. Darwis, and A. R. Manga, "Analysis of ensemble machine learning classification comparison on the skin cancer MNIST dataset," *Computer Science and Information Technologies*, vol. 5, no. 3, pp. 235–242, 2024, doi: 10.11591/csit.v5i3.pp235-242.

[16] C. Sirocchi, A. Bogliolo, and S. Montagna, "Medical-informed machine learning: integrating prior knowledge into medical decision systems," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. Suppl 4, Dec. 2024, doi: 10.1186/s12911-024-02582-4.

[17] Z. Zulham, E. E. Putri, and B. S. Hasugian, "Pattern Analysis of Drug Procurement System With FP-Growth Algorithm," *Jurnal Online Informatika*, p., 2022, doi: 10.15575/join.v7i1.841.

[18] B. Anwar, A. Ambiyar, and F. Fadhilah, "Application of the FP-Growth Method to Determine Drug Sales Patterns," *Sinkron*, vol. 8, no. 1, pp. 405–414, Jan. 2023, doi: 10.33395/sinkron.v8i1.12004.

[19] J. Nugraha and C. Yustia Purnamawati, "Application of the Association Method Using FP-Growth Algorithm to Find Pattern Of Medicine Purchasing Transactions at Pharmacy," *Journal of Applied Research In Computer Science and Information Systems*, vol. 1, no. 2, pp. 48–53, Nov. 2023, doi: 10.61098/jarcis.v1i2.50.

[20] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/4832864.

[21] Herman, H. Darwis, Nurfauziyah, R. Puspitasari, D. Widyawati, and A. Faradibah, "Comparative Analysis of Anxiety Disorder Classification Using Algorithm Naïve Bayes, Decision Tree and K-NN," in *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2025, pp. 1–6. doi: 10.1109/IMCOM64595.2025.10857485.

[22] H. Essalmi and A. El Affar, "Dynamic Algorithm for Mining Relevant Association Rules via Meta-Patterns and Refinement-Based Measures," *Information (Switzerland)*, vol. 16, no. 6, Jun. 2025, doi: 10.3390/info16060438.

[23] M. Rosenblatt, L. Tejavibulya, R. Jiang, S. Noble, and D. Scheinost, "Data leakage inflates prediction performance in connectome-based machine learning models," *Nat. Commun.*, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-46150-w.

[24] Z. Wu and G. Fang, "The Research on the Improvement of FP-growth Algorithm," *Artificial Intelligence Technology Research*, vol. 2, no. 1, Mar. 2024, doi: 10.18686/aitr.v2i1.3860.

[25] C. Fernandez-Basso, M. D. Ruiz, and M. J. Martin-Bautista, "New Spark solutions for distributed frequent itemset and association rule mining algorithms," *Cluster Comput.*, vol. 27, no. 2, pp. 1217–1234, Apr. 2024, doi: 10.1007/s10586-023-04014-w.

[26] A. Ahmadipour and A. Sarafinejad, "Investigation of Drug Interactions through Analysis of Prescribed Medications Association Rules Using the FP -growth Algorithm," *Journal of Health and Biomedical Informatics*, vol. 11, no. 2, pp. 166–175, Sep. 2024, doi: 10.34172/jhbmi.2024.22.

[27] W. Yang *et al.*, “Breaking the Top-K Barrier: Advancing Top-K Ranking Metrics Optimization in Recommender Systems,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2025, pp. 3542–3552. doi: 10.1145/3711896.3736866.

[28] H. Darwis, F. A. Syahrir, and L. N. Hayati, “A Hybrid Movie Recommendation System to Address Data Sparsity Using Genre-Based K-Means and Neural Collaborative Filtering,” *ILKOM Jurnal Ilmiah*, vol. 17, no. 2, pp. 203–212, Sep. 2025, doi: 10.33096/ilkom.v17i2.2868.203-212.

[29] H. Lahuddin, M. R. A. Muliawan, K. Takemoto, H. Darwis, S. R. Jabir, and R. Adawiyah, “Cryptocurrency Prices Forecasting Using LSTM, CNN, Transformer, TCN, and Hybrid Model: A Deep Learning Approach,” in *2025 9th International Conference On Electrical, Electronics And Information Engineering (ICEEIE)*, 2025, pp. 1–6. doi: 10.1109/ICEEIE66203.2025.11252474.

[30] B. Drammeh and H. Li, “Enhancing neural collaborative filtering using hybrid feature selection for recommendation,” *PeerJ Comput. Sci.*, vol. 9, 2023, doi: 10.7717/peerj-cs.1456.

[31] X. Meng, “Cross-domain information fusion and personalized recommendation in artificial intelligence recommendation system based on mathematical matrix decomposition,” *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-57240-6.

[32] W. Y. Tan, Q. Gao, R. W. Oei, W. Hsu, M. L. Lee, and N. C. Tan, “Diabetes medication recommendation system using patient similarity analytics,” *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-24494-x.

[33] A. Sae-Ang, S. Chairat, N. Tansuebchueasai, O. Fumaneeshoat, T. Ingviya, and S. Chaichulee, “Drug Recommendation from Diagnosis Codes: Classification vs. Collaborative Filtering Approaches,” *Int. J. Environ. Res. Public Health*, vol. 20, no. 1, Jan. 2023, doi: 10.3390/ijerph20010309.

[34] P. Cui, B. Yin, and B. Xu, “The application of social recommendation algorithm integrating attention model in movie recommendation,” *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-43511-1.

[35] R. P. Tapio, “The Role of Data Assumptions in Selecting Between Parametric and Nonparametric Tests,” *Asian Journal of Probability and Statistics*, vol. 27, no. 11, pp. 127–135, Nov. 2025, doi: 10.9734/ajpas/2025/v27i11830.

[36] Z. Kala, “Global Sensitivity Analysis of Structural Reliability Using Cliff Delta,” *Mathematics*, vol. 12, no. 13, Jul. 2024, doi: 10.3390/math12132129.