Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 8, No. 3, November 2025, pp. 536 – 544
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Discovering Prescription Patterns in Type 2 Diabetes Based on Demographic Attributes Using Association Rules

[1*]Putri Yani, [2]Maulida Hikmah, [3]Deni Mahdiana
[1,2,3]Master of Computer Science, Faculty of Information Technology, Budi Luhur University, Indonesia
Email: [1]2311601450@student.budiluhur.ac.id, [2]2311601484@student.budiluhur.ac.id, [3]deni.mahdiana@budiluhur.ac.id

Article Info
Article history:
Received Aug 03rd, 2025
Revised Sep 13th, 2025
Accepted Sep 25th, 2025

Keyword:
Apriori
Data Mining
Demographic Analysis
Diabetes Mellitus
Prescription Patterns

ABSTRACT
Type 2 diabetes mellitus (T2DM) is a chronic disease that requires effective long-term therapeutic management. Appropriate and continuous treatment is crucial for preventing complications and improving patients' quality of life. In clinical practice, prescription patterns vary significantly and are influenced by demographic and clinical characteristics. This study aimed to analyze prescription patterns of T2DM patients based on demographic and clinical attributes, and to identify frequently co-prescribed drug combinations using the Apriori algorithm. A total of 3,500 prescription records were obtained from RSUD H. Damanhuri Barabai. The analysis was conducted in two stages: (1) association between demographic factors (age, gender, blood pressure) and prescribed drugs, and (2) association among drugs regardless of patient demographics. With a minimum support of 3%, confidence thresholds of 60% and 35%, and a lift greater than 1.5, fifteen valid rules were identified in the demographic-to-drug analysis, and nine rules in the drug combination analysis. Strong patterns were observed, such as the prescription of Empagliflozin and Insulin Degludec for hypertensive patients aged 40–49, and the co-prescription of Acarbose and Glimepiride. These findings demonstrated that the Apriori algorithm was effective in identifying meaningful prescription patterns. Beyond methodological contributions, the results provide practical value for hospitals by supporting pharmacy managers in drug procurement planning, optimizing stock management, and designing distribution strategies that anticipate patient needs based on prescription trends.

Corresponding Author:
Putri Yani,
Master of Computer Science, Faculty of Information Technology
Budi Luhur University
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, DKI Jakarta 12260, Indonesia
Email: 2311601450@student.budiluhur.ac.id
DOI: http://dx.doi.org/10.24014/ijaidm.v8i3.38082

1. INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a chronic metabolic disorder characterized by elevated blood glucose levels due to impaired insulin secretion or action, which may lead to serious complications in the eyes, kidneys, nerves, and heart [1]. Increasing risk factors such as high-calorie diets, low fiber intake, and physical inactivity have contributed to the global prevalence. The International Diabetes Federation (IDF) reported that 537 million adults were living with diabetes in 2021, and this number is projected to reach 783 million by 2045 [2]. In Indonesia, the prevalence increased from 6.9% in 2013 to 8.5% in 2018 [3], , with hypertension and obesity further aggravating complications [4]. In clinical practice, prescriptions for T2DM patients often vary even when patients have similar conditions. A study at Sidoarjo Hospital found that most patients received oral antidiabetic drugs (OAD) as monotherapy, while others received combinations of OAD and insulin [5]. Another study also reported variations in the use of metformin, glimepiride, and combinations such as metformin–glimepiride–acarbose in Indonesian healthcare facilities [6].

The long-term impact of T2DM requires comprehensive management, including lifestyle interventions and pharmacological therapy [3]. In practice, pharmacological therapy is often the primary treatment, particularly for patients who are unable to control their glucose levels through lifestyle modifications alone. However, prescription practices show considerable variability, influenced by patients' clinical conditions, therapeutic responses, comorbidities, and institutional policies [7].

The wide range of pharmacological options indicates that no single therapy suits all patients. Metformin is generally used as the first-line therapy due to its effectiveness, while alternatives such as sulfonylureas, Dipeptidyl Peptidase-4 (DPP-4) inhibitors, alpha-glucosidase inhibitors, or Sodium Glucose Co Transporter-2 (SGLT-2) inhibitors are prescribed if metformin is unsuitable [6]. If HbA1c targets are not achieved, combination therapy is recommended, while insulin is administered in cases of severe hyperglycemia or advanced complications [8].

Variations in prescription practices pose challenges in establishing consistent and effective treatment patterns. The absence of structured information on commonly prescribed drug combinations also hampers the optimization of diabetes management [9]. Moreover, factors such as age, disease severity, and hospital pharmacy policies influence drug selection [10]. Therefore, data-driven analysis is required to identify prescription patterns and commonly used drug combinations in clinical practice.

Previous studies have demonstrated the effectiveness of the Apriori algorithm in analyzing prescription data. Saputra and Sibarani [11] applied it to pharmacy sales data and identified high-confidence association rules useful for stock planning, while Nola et al. [12] used it on hospital medical records to generate association rules relevant to healthcare policy recommendations. The Apriori algorithm identifies frequent itemsets and generates association rules based on support and confidence metrics [13]. Besides Apriori, the Frequent Pattern Growth (FP-Growth) algorithm is also widely used; however, testing in this study showed that Apriori produced rules of identical quality to FP-Growth while achieving faster execution times and simpler interpretation. This makes Apriori more suitable for clinical analysis and validation by medical professionals. Its application is expected to uncover trends and effective drug combinations, thereby supporting clinical decision-making and hospital inventory planning [14]. This study is also expected to provide practical benefits for hospital management and pharmacy inventory units in understanding prescription patterns for T2DM patients. Such information can serve as a basis for drug procurement planning, stock management, and distribution strategies that are more efficient and targeted, thereby strengthening hospital pharmaceutical logistics policies in anticipating future drug needs according to prescription trends.

2. RESEARCH METHOD

The research method employed in this study adopted the Cross Industry Standard Process for Data Mining (CRISP–DM) framework, which consists of six stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The overall research procedure is illustrated in Figure 1.

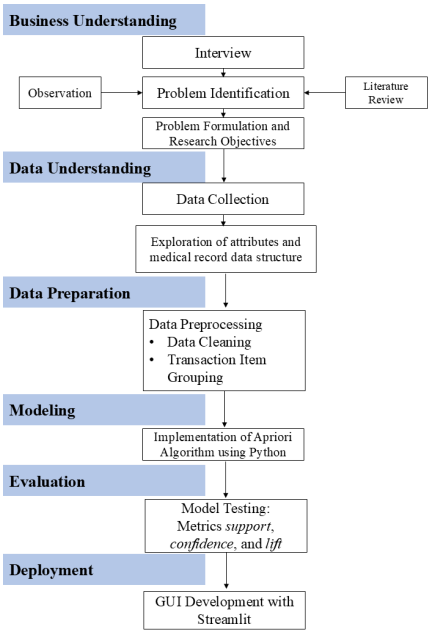

Figure 1. Research Procedure Using the CRISP–DM Methodology

2.1 Business Understanding

T2DM is a chronic disease with a high prevalence that requires long-term therapeutic management. Variations in prescription practices often occur due to differences in patient conditions and clinical considerations, making it difficult to identify consistent prescribing patterns. In the CRISP–DM framework, the Business Understanding phase is intended to align data mining objectives with business or organizational goals, ensuring that the analysis is relevant and problem-oriented [15,16]. Therefore, this study aims to discover prescription patterns based on patients' demographic and clinical attributes using the Apriori algorithm, as well as to identify the most frequently co-prescribed drug combinations. The findings are expected to support data-driven clinical decision-making and hospital pharmacy inventory planning, in line with the emphasis of CRISP–DM that outcomes must contribute directly to operational needs [17].

2.2 Data Understanding

The data understanding stage aimed to provide an overview of the research dataset. The dataset was obtained from the medical records of T2DM patients at RSUD H. Damanhuri Barabai during the period January 2023–December 2024, consisting of 3,500 prescription transactions. The initial extraction from the medical record system included patient identity and medical history. From this source, a research dataset was derived, including demographic attributes (gender, age), clinical attributes (blood pressure), and prescribed drugs.

To ensure consistency, drug names initially recorded by brand were standardized into their generic forms. Standardizing drug names is an important step to reduce duplication and maintain data quality [18]. The dataset was then converted into binary format using the one-hot encoding technique, where a value of 1 indicates that a drug was prescribed and 0 indicates otherwise. One-hot encoding is a common method in data mining for representing categorical attributes in numerical form so that they can be processed by analytical algorithms [19]. The final output of this stage was an integrated dataset prepared for the data preparation phase and subsequent analysis using the Apriori algorithm, in line with the CRISP–DM principle that emphasizes initial data exploration to understand characteristics and potential quality issues [17].

2.3 Data Preparation

The data preparation stage aimed to transform the integrated dataset into a format suitable for modeling with the Apriori algorithm. Several steps were carried out sequentially, starting with data merging, in which demographic (gender, age), clinical (blood pressure), and prescription data were integrated using patient ID as the primary key. This process produced a single dataset containing all relevant attributes, in line with best practices for integrating heterogeneous healthcare data [20]. Next, data cleaning was performed to handle incomplete entries and check for inconsistencies in demographic and clinical information. Data cleaning is an essential step in ensuring data quality and reliability before modeling [20]. Attribute selection was then applied by focusing on gender (male, female), age groups (<30, 30–39, 40–49, 50–59, 60–69, ≥70 years), blood pressure categories (normal, elevated, high), and prescriptions related to diabetes management and comorbidities (e.g., Metformin, Acarbose, Glimepiride, Sitagliptin, Vildagliptin, Canagliflozin, Dulaglutide, Empagliflozin, Liraglutide, and various types of insulin). Attribute selection improves prediction performance, reduces complexity, and enhances interpretability [21].

To ensure consistency, prescriptions initially recorded using brand names were standardized into generic names. For instance, Metformin XR 750 mg and Metformin 500 mg were grouped as Metformin, Tresiba was standardized as Insulin Degludec, and Novorapid Flexpen as Insulin Aspart. This process aligns with recommended informatics frameworks for standardized medication data collection and analysis [22]. Finally, the dataset was transformed into a binary transactional format using one-hot encoding, where "1" indicates that a drug was prescribed and "0" indicates otherwise. One-hot encoding is a widely used preprocessing technique to represent categorical attributes in numerical form for data mining algorithms [19]. An example of the transformed dataset is presented in Table 1.

Table 1. Example of Transformed Dataset

Number	Gender	Age	BP	Acarbose	Canagliflozin	Dulaglutide	…	Vildagliptin
1	F	50-59	Normal BP	0	0	1	..	0
2	F	30-39	Normal BP	0	0	0	..	0
..	..	..	..	..	..	..	..	..
3500	M	60-69	High BP	0	0	0	..	0

With this format, each row represented a patient along with demographic, clinical, and prescription attributes, enabling the Apriori algorithm to identify frequent itemsets and generate association rules.

2.4 Modeling

The modeling stage aimed to analyze prescription patterns in T2DM patients and identify drug combinations that are frequently prescribed together. The approach employed was association rule mining using the Apriori algorithm, which is widely recognized for its efficiency in identifying relationships among items in binary datasets [23,24].

The research dataset included demographic attributes (gender: male, female; age: <30, 30–39, 40–49, 50–59, 60–69, ≥70 years), clinical attributes (blood pressure: normal, elevated, high), and binary variables representing antidiabetic and comorbidity-related drugs such as Metformin, Acarbose, Glimepiride, Sitagliptin, Vildagliptin, Repaglinide, Canagliflozin, Dulaglutide, Empagliflozin, Liraglutide, and various types of insulin. Each row represented an individual patient, with a combination of demographic, clinical, and prescription attributes. A value of "1" indicated that a drug was prescribed, and "0" indicated otherwise.

The analysis was conducted through two main approaches. First, Demographic → Drug, in which antecedents consisted of patient characteristics (gender, age, blood pressure) and consequents were the prescribed drugs. This analysis aimed to identify correlations between patient profiles and drug utilization. Second, Drug → Drug, in which each drug was treated as an item, and Apriori was applied to identify co-prescription patterns. Association rules were generated from frequent itemsets that met the thresholds of support, confidence, and lift, which are the standard measures for evaluating association rule quality [25]. Rules were then assessed for both statistical validity and clinical relevance, as recommended in medical applications of association rule mining [26].

2.5 Evaluation

The evaluation stage was conducted to ensure that the Apriori algorithm generated association rules that were both statistically significant and clinically relevant. Statistical validation was carried out using three main metrics:

1. Support, which measures the proportion of transactions that contain a particular itemset. It is calculated as the ratio between the number of transactions containing the itemset X and the total number of transactions, as shown in equation (1) [18].

$$\text{Support}\ (X) = \frac{\text{Number of transactions containing (X)}}{\text{Total number of transactions}} \tag{1}$$

2. Confidence, which measures the probability that the consequent Y appears in transactions that contain the antecedent X. It is defined as the ratio of support for X ∪ Y to the support of X, as shown in equation (2) [27].

$$\text{Confidence}\ (X \rightarrow Y) = \frac{\text{Support (X} \cap \text{Y)}}{\text{Support (X)}} \tag{2}$$

3. Lift, which evaluates the strength of the relationship between antecedent X and consequent Y compared to random chance. It is expressed as the ratio of support for X ∪ Y to the product of support for X and support for Y, as shown in equation (3) [28].

$$\text{Lift}\ (X \rightarrow Y) = \frac{\text{Support (X} \cup \text{Y)}}{\text{Support (X)} \times \text{Support (Y)}} \tag{3}$$

This evaluation was applied to rules linking patient characteristics with prescribed drugs (Demographic → Drug) as well as rules identifying drug combinations frequently prescribed together (Drug → Drug). Threshold values for these metrics were determined through preliminary experiments to filter rules with meaningful associations.

In addition to statistical validation, clinical validation was performed through discussions with medical practitioners and comparison with established guidelines from the American Diabetes Association (ADA) [29], the Indonesian Society of Endocrinology (PERKENI) [8], and related literature. The resulting rules were presented in the form of tables, heatmaps, network graphs, and scatter plots to facilitate interpretation. These visualizations supported the identification of prescribing patterns consistent with clinical practice and provided additional insights for healthcare professionals in making more data-driven therapeutic decisions.

2.6 Deployment

The deployment stage aimed to implement the research findings into a practical tool. A simple prototype was developed using Streamlit, which allows users to interactively explore association rules generated by the Apriori algorithm. Through this interface, healthcare practitioners can identify prescription

patterns based on patient demographics or view frequently co-prescribed drugs. This prototype demonstrates the potential application of the proposed model in supporting hospital pharmacy management and clinical decision-making.

The use of Streamlit for deploying predictive models in healthcare has been recognized as an effective approach to bridge the gap between research and practical implementation, offering clinicians access to interactive decision support tools without requiring advanced programming skills [30].

3. RESULTS AND ANALYSIS

3.1 Dataset Description

The research dataset consisted of 3,500 prescription records of T2DM patients obtained from RSUD H. Damanhuri Barabai during the period of January 2023–December 2024. The integrated dataset included demographic attributes (gender, age), clinical attributes (blood pressure), and prescribed drugs.

Patient age was grouped into six categories: <30 years, 30–39 years, 40–49 years, 50–59 years, 60–69 years, and ≥70 years. Blood pressure (BP) was categorized into Normal BP, Elevated BP, and High BP according to the American Heart Association (AHA) standards. Prescriptions were standardized into generic names, covering antidiabetic and comorbidity-related drugs such as Metformin, Acarbose, Glimepiride, Sitagliptin, Vildagliptin, Repaglinide, Canagliflozin, Dulaglutide, Empagliflozin, Liraglutide, and various types of insulin, as summarized in Table 2.

Table 2. Dataset Attributes and Categories

Number	Attribute	Categories/ Types
1	Gender	Male, Female
2	Age	<30, 30–39, 40–49, 50–59, 60–69, ≥70 years
3	Blood Pressure	Normal BP, Elevated BP, High BP
4	Drug	Metformin, Acarbose, Glimepiride, Sitagliptin, Vildagliptin, Repaglinide, Canagliflozin, Dulaglutide, Empagliflozin, Liraglutide, Insulin (various types)

The final dataset was then converted into binary format using the one-hot encoding technique, where each row represented a patient with demographic, clinical, and prescription attributes. An example of the converted dataset is shown in Table 3.

Table 3. Example of One-Hot Encoding Conversion

Number	Gender	Age	BP	Acarbose	Canagliflozin	Dulaglutide	…	Vildagliptin
1	F	50-59	Normal BP	0	0	1	..	0
2	F	30-39	Normal BP	0	0	0	..	0
3	M	60-69	High BP	0	0	0	..	0
4	M	30-39	High BP	0				
5	F	50-59	High BP	0	1	1	…	0
..	..	…	…	…	…	…	…	…
3500	M	Under 30	Elevated BP	0	0	1	…	0

With this format, the dataset was ready for analysis using the Apriori algorithm to generate association rules.

3.2 Result of Demographic → Drug Analysis

The first stage of analysis explored the associations between patient demographic and clinical characteristics and the prescribed drugs. As described in the data pre-processing stage, prescription data were standardized into generic names, merged with demographic and clinical attributes, and transformed into a binary format using the one-hot encoding technique. This ensured that the generated association rules could be systematically traced back to the integrated pre-processed dataset (see Table 3). Based on this dataset, the Apriori algorithm was executed with minimum support of 3%, confidence of 60%, and lift > 1.5, resulting in a total of 15 valid association rules. Several representative rules with the highest support, confidence, and lift values are presented in Table 4 as examples of the discovered patterns.

From a statistical perspective, lift values indicate the strength of associations beyond random chance. Rule 5 (Female ≥70 years, Elevated BP → Insulin Detemir) shows the highest lift (9.080), meaning that the likelihood of prescribing Insulin Detemir in this group is nine times higher than by random occurrence, demonstrating a very strong association. Rule 1 (Empagliflozin) also indicates a strong relationship (lift 6.137), while Rule 2 (Insulin Aspart) has a weaker but still meaningful association (lift 3.202). The support values ranging from 0.030–0.045 suggest that these patterns, while not dominant across the entire dataset, consistently emerge within specific demographic subgroups.

Clinically, these findings align with established diabetes treatment guidelines. The frequent prescription of Metformin (Rule 4) is consistent with its role as the first-line therapy for type 2 diabetes [PERKENI 2021; ADA 2023]. The strong association between Empagliflozin (Rule 1) and female patients aged 40–49 with hypertension reflects the cardioprotective benefits of Sodium Glucose Co-Transporter-2 (SGLT2) inhibitors, recommended for patients with cardiovascular risk factors. The prescription of Insulin Aspart and Insulin Degludec (Rules 2 and 3) for patients aged 40–49 suggests the need for tighter glycemic control involving both prandial and basal regimens. Meanwhile, the very strong association observed in Rule 5 highlights the clinical practice of initiating long-acting insulin such as Insulin Detemir in elderly patients with poor glycemic control and comorbid hypertension.

From a practical perspective, these results emphasize that age and blood pressure are primary determinants in prescribing decisions, while gender further differentiates treatment tendencies among specific groups. For hospital pharmacy management, the relatively higher support values for Metformin and Insulin Detemir indicate the need to prioritize these drugs in inventory planning to anticipate demand from patients with similar demographic and clinical profiles. Thus, association rule mining not only reveals statistically significant patterns but also clinically meaningful insights that can support pharmacy logistics planning.

Table 4. Selected Association Rules (Demographic → Drug)

Number	Antecedent (Demographics)	Consequent (Drug)	Support	Confidence	Lift
1	Female, 40–49, High BP	Empagliflozin	0.030	0.684	6.137
2	Female, 40–49, High BP	Insulin Aspart	0.033	0.755	3.202
3	Female, 40–49, High BP	Insulin Degludec	0.033	0.735	4.054
4	Male, 60–69, High BP	Metformin	0.045	0.799	3.734
5	Female, ≥70, Elevated BP	Insulin Detemir	0.045	0.739	9.080

3.3 Result of Drug → Drug Analysis

The second stage of analysis focused on identifying associations among prescribed drugs to uncover frequent co-prescription patterns independent of patient demographics. As described in the data pre-processing stage, drug names were standardized into generic forms and transformed into a binary transactional format, allowing the Apriori algorithm to detect meaningful combinations. With minimum support of 3%, confidence threshold of 35%, and lift > 1.5, a total of nine valid association rules were generated, all of which are presented in Table 5.

Table 5. Association Rules of Drug-to-Drug Combinations

Number	Antecedent (Drug)	Consequent (Drug)	Support	Confidence	Lift
1	Acarbose	Glimepiride	0.043	1	7.642
2	Canagliflozin	Insulin Glargine	0.078	0.508	1.583
3	Canagliflozin	Repaglinide	0.055	0.356	2.284
4	Repaglinide	Canagliflozin	0.055	0.350	2.284
5	Empagliflozin	Insulin Degludec	0.040	0.359	1.979
6	Glipizide	Insulin Lispro	0.039	0.368	2.484
7	Insulin Degludec	Insulin Aspart	0.083	0.457	1.937
8	Insulin Aspart	Insulin Degludec	0.083	0.352	1.937
9	Repaglinide	Insulin Glargine	0.078	0.501	1.560

From a statistical perspective, Rule 1 (Acarbose → Glimepiride) stands out with the highest lift value (7.642), indicating that these two drugs are very strongly associated and frequently prescribed together. The support value of 0.043 shows that this combination, while not dominant across all prescriptions, consistently appears within the dataset. Rules 7 and 8 (Insulin Degludec ↔ Insulin Aspart) demonstrate reciprocal associations with support 0.083, confirming the frequent use of basal–bolus insulin regimens. Meanwhile, Rule 6 (Glipizide → Insulin Lispro) has a moderate support but relatively high lift (2.484), highlighting its strength as a non-random combination.

Clinically, these results align with current treatment guidelines. The combination of Acarbose and Glimepiride (Rule 1) reflects the practice of combining alpha-glucosidase inhibitors with sulfonylureas to enhance glycemic control. The frequent pairing of Insulin Degludec and Insulin Aspart (Rules 7 and 8) illustrates the use of basal–bolus therapy, recommended for patients requiring intensive glycemic control [PERKENI 2021; ADA 2023]. The associations involving SGLT2 inhibitors (Empagliflozin, Canagliflozin) with insulin or secretagogues such as Repaglinide (Rules 2–5, 9) highlight their role in patients with poor control on oral therapy and additional cardiovascular risk factors.

From a practical perspective, these findings indicate that certain drug combinations, particularly insulin regimens and oral drug pairings, dominate prescribing patterns in T2DM management. For hospital pharmacy management, the relatively high support values of Insulin Aspart and Insulin Degludec suggest the need for adequate stock of both basal and prandial insulin analogs. Meanwhile, the presence of combinations

involving SGLT2 inhibitors underscores the importance of ensuring their availability in line with the growing clinical adoption of these medications. These insights reinforce the value of association rule mining in uncovering statistically robust and clinically meaningful prescription patterns, while also guiding strategic inventory planning.

3.4 Discussion

The results of this study show that demographic and clinical factors significantly influence prescribing patterns in patients with T2DM. Rules derived from the demographic-to-drug analysis revealed that age and blood pressure were the strongest determinants, with gender acting as a secondary factor. For example, female patients aged 40–49 years with high blood pressure were more likely to receive Empagliflozin, Insulin Aspart, and Insulin Degludec, while male patients aged 60–69 years with high blood pressure had a strong association with Metformin prescriptions. These findings are consistent with the recommendations of the ADA [29] and the PERKENI [8], which emphasize the need for therapy tailored to patient profiles, cardiovascular risk, and comorbid conditions.

The drug-to-drug analysis highlighted several important co-prescription patterns. The strongest association was found between Acarbose and Glimepiride, a combination previously reported as effective in improving glycemic control in dual oral therapy. Likewise, the reciprocal association between Insulin Degludec and Insulin Aspart reflects standard clinical practice of combining basal and prandial insulin in patients with advanced T2DM or poor glycemic control. Modern therapeutic patterns, such as Empagliflozin with Insulin Degludec and Canagliflozin with Repaglinide, were also identified, in line with current clinical guidelines that recommend combining SGLT2 inhibitors with insulin or secretagogues for patients at risk of cardiovascular complications.

In interpreting these findings, demographic and clinical attributes such as gender, age, and blood pressure play a central role in shaping prescribing behavior. Gender differences have biological and psychosocial impacts that influence disease progression, complications, and therapeutic response. Kautzky-Willer and Harreiter (2017) [31] emphasized that men and women show differences in risk factors, therapy preferences, drug side effects, and adherence to lifestyle or pharmacological interventions. Therefore, considering gender differences in therapy is essential to improve the quality of care and clinical outcomes. Patient age was grouped into six categories for analytical purposes. This is consistent with the recommendations of the IDF [2], which highlights the importance of considering older age in diabetes management, as elderly patients often require more intensive regimens, including insulin. Blood pressure status was also a critical determinant, given that hypertension is the most common comorbidity in T2DM, and international guidelines from the World Health Organization [32] and the Joint National Committee [33] recommend maintaining blood pressure below 130/80 mmHg to prevent cardiovascular complications.

The emergence of strong associations between insulin combinations, particularly Insulin Degludec and Insulin Aspart, can be explained by the clinical need for tighter glycemic control in patients who fail oral therapy. Elderly patients or those with multiple comorbidities often cannot achieve adequate control with monotherapy or dual oral therapy, necessitating the initiation of basal–bolus insulin regimens. This finding is consistent with international guidelines such as ADA (2023) and PERKENI (2021), which recommend basal–prandial insulin combinations in advanced T2DM or cases with severe hyperglycemia.

Overall, the findings strengthen the clinical validity of the Apriori algorithm in identifying prescribing patterns consistent with medical guidelines and real-world practice. Moreover, the discovered association rules provide valuable insights for healthcare professionals to design more personalized, data-driven treatment strategies and to support hospital pharmacy inventory planning.

4. CONCLUSION

This study applied the Apriori algorithm to analyze prescribing patterns in patients with T2DM. The results showed that age, blood pressure, and gender significantly influenced prescribing behavior, while the drug-to-drug analysis revealed key co-prescription patterns such as Acarbose–Glimepiride and Insulin Degludec–Insulin Aspart, as well as modern combinations with SGLT2 inhibitors. The strength of this study lies in the use of real-world hospital data, systematic preprocessing steps, and the interpretability of the Apriori algorithm, which ensured the generation of clinically meaningful rules. However, some limitations should be acknowledged, including the use of data from a single hospital, the absence of key clinical variables such as HbA1c or Body Mass Index (BMI), and a restricted study period of two years.

Overall, the findings confirm the clinical validity of association rule mining in supporting personalized therapy design and hospital pharmacy inventory planning. Future research is recommended to include additional clinical parameters, expand to multi-center datasets, and explore alternative algorithms such as Eclat to strengthen the generalizability of the results.

REFERENCES

[1] Johnson EL, Feldman H, Butts A, Chamberlain J, Collins B, Doyle-Delgado K, et al. Standards of medical care in diabetes—2020 abridged for primary care providers. Clin Diabetes. 2020;38(1):10–38.

[2] IDF. International Diabetes Federation. Diabetes Research and Clinical Practice. 2021.

[3] Kementerian Kesehatan RI. Laporan Nasional Riskesdas 2018 [Internet]. Lembaga Penerbit Balitbangkes. 2018. p. hal 156. Available from: https://repository.badankebijakan.kemkes.go.id/id/eprint/3514/1/Laporan Riskesdas 2018 Nasional.pdf

[4] Adi S. Pengelolaan Dan Pencegahan Diabetes Melitus Tipe 2 Dewasa di Indonesia. PB Perkeni. 2019;133.

[5] Susanto Y, Bangun Lestari S, Purwati E. Pola Peresepan Obat Antidiabetik Pada Pasien Diabetes Melitus Tipe 2 Di Poli Penyakit Dalam Rs "X" Sidoarjo Periode Oktober- Desember 2019. 2020;1(2):69–80.

[6] Artini KS, Ajeng T, Saifana CS. Gambaran Penggunaan Antidiabetes Oral Pada Pasien Diabetes Mellitus Tipe 2. SIKesNas. 2022;2045:333–40.

[7] Jiang L, Chang J, Wang Y, Ma X. Pharmacological therapies for type 2 diabetes: a review of clinical trials and real-world studies. J Clin Med. 2020;9(7):1–20.

[8] Perkumpulan Endokrinologi Indonesia. Pedoman Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 di Indonesia 2021. PB Perkeni; 2021.

[9] Astuti IW, Suryadi T. Faktor-faktor yang mempengaruhi pola resep obat pada pasien diabetes melitus. J Farm Klin Indones. 2020;9(1):25–32.

[10] Mohan V, Deepa M, Pradeepa R, Anjana RM. Therapeutic options for diabetes in low- and middle-income countries. Lancet Diabetes Endocrinol. 2019;7(1):1–12.

[11] Saputra R, Sibarani AJP. Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat. JATISI (Jurnal Tek Inform dan Sist Informasi). 2020;7(2):262–76.

[12] Nola R, Suswaini E, Pebriadi W. Penerapan Association Rule Menggunakan Algoritma Apriori Pada Poliklinik Penyakit Dalam (Studi Kasus: Rumah Sakit Umum Daerah Bintan). J Sains dan Inform. 2021;7(2):222–30.

[13] Kumar R, Singh R, Jain A. Apriori Algorithm for Finding Frequent Itemsets in Big Data. J Comput Sci. 2018;14(2):120–7.

[14] Tandel KR, Patel A, Thakkar S. Use of association rule learning in clinical prescription patterns: A review. Int J Heal Med Sci. 2020;10(2):67–74.

[15] Azevedo A, Santos MF. Data Mining and CRISP-DM Methodology. In: Big Data Processing and Applications. Springer, Cham; 2018. p. 245–64.

[16] Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, et al. CRISP-DM 1.0 Step-by-Step Data Mining Guide: Revisited for Big Data Applications. SPSS; 2020.

[17] Kastouni MZ, Ait Lahcen A. Big data analytics in telecommunications: Governance, architecture and use cases. J King Saud Univ - Comput Inf Sci. 2022;34(6):2758–70.

[18] Han J, Pei J, Kamber M. Data Mining: Concepts and Techniques. 3rd Edition. Morgan Kaufmann; 2019.

[19] García S, Luengo J, Herrera F. Data Preprocessing in Data Mining. Vol. 72. 2015.

[20] Rahm E, Do H. Data cleaning: Problems and current approaches. IEEE Data Eng Bull [Internet]. 2000;23(4):3–13. Available from: http://wwwiti.cs.uni-magdeburg.de/iti_db/lehre/dw/paper/data_cleaning.pdf%5Cnpapers2://publication/uuid/17B58056-3A7F-4184-8E8B-0E4D82EFEA1A%5Cnhttp://dc-pubs.dbs.uni-leipzig.de/files/Rahm2000DataCleaningProblemsand.pdf

[21] Guyon I. An Introduction of Variabel and Feature Selection. CrossRef List Deleted DOIs. 2015;1(April).

[22] Richesson RL. An informatics framework for the standardized collection and analysis of medication data in networked research. J Biomed Inform [Internet]. 2014;52:4–10. Available from: http://dx.doi.org/10.1016/j.jbi.2014.01.002

[23] Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules in Large Databases. In: Very Large Data Bases Conference [Internet]. 1994. Available from: https://api.semanticscholar.org/CorpusID:3131928

[24] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. SIGMOD Rec [Internet]. 1993 Jun;22(2):207–216. Available from: https://doi.org/10.1145/170036.170072

[25] Tan PN, Steinbach M, Kumar V. Introduction to Data Mining. 2nd ed. Pearson; 2018.

[26] Altaf W, Shahbaz M, Guergachi A. Applications of association rule mining in health informatics: a survey. Artif Intell Rev [Internet]. 2017;47(3):313–40. Available from: https://doi.org/10.1007/s10462-016-9483-9

[27] Mahmood S, Shahbaz M, Guergachi A. Negative and positive association rules mining from text using frequent and infrequent itemsets. Sci World J. 2014;2014.

[28] Hsun-Lu P, Keng JL, Tsai FM, Lu PH, Kou CY. An Apriori Algorithm-Based Association Rule Analysis to.pdf. 2021.

[29] American Diabetes Association. Standards of Medical Care in Diabetes—2019. Diabetes Care. 2019;42(Suppl 1):S1--S193.

[30] Wallard L. Model deployment of a predictive machine learning model for clinical decision support. Bachelor Thesis, Leiden Inst Adv Comput Sci. 2023;

[31] Kautzky-Willer A, Harreiter J. Sex and gender differences in therapy of type 2 diabetes. Diabetes Res Clin Pract [Internet]. 2017;131(July):230–41. Available from: http://dx.doi.org/10.1016/j.diabres.2017.07.012

[32] World Health Organization. Global report on diabetes. Geneva: World Health Organization; 2020.

[33] Joint National Committee. The seventh report of the joint national committee on the prevention, detection, evaluation and treatment of high blood pressure: The weapons are ready. 2004.

BIBLIOGRAPHY OF AUTHORS

Putri Yani, I obtained my Bachelor's degree in Computer Science Education from Universitas Lambung Mangkurat and my Master's degree in Computer Science from Universitas Budi Luhur. My research interests include data mining, artificial intelligence, and healthcare informatics.

Maulida Hikmah, I earned my Bachelor's degree in Informatics Engineering from Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin and continued my Master's degree in Computer Science at Universitas Budi Luhur. My research interests are in machine learning, big data analytics, and clinical decision support systems.

Deni Mahdiana, I am a lecturer at the Faculty of Information Technology, Universitas Budi Luhur, and currently serve as Deputy Rector for Non-Academic Affairs at Universitas Budi Luhur. My research focuses on database development, database management, software development, data mining, and knowledge discovery.