**Performance Assessment**

OFM4 — OFM4 Task 3: Association Rules and Lift Analysis

Bader Ale

Department of Information Technology, Western Governors University

D212 Data Mining II

2024

Table of Contents

[Part I: Research Question 3](#_Toc185526608)

[Part II: Method Basket Justification 3](#_Toc185526609)

[Part III: Data Preparation and Analysis 4](#_Toc185526610)

[Part IV: Data Summary and Implications 7](#_Toc185526611)

# Part I: Research Question

The purpose of this report is to explore the following research question: "How can Market Basket Analysis be used to identify meaningful prescription associations that support strategic decision-making for reducing patient readmissions and improving hospital cost-effectiveness?" By applying Market Basket Analysis to historical patient prescription data, the goal is to uncover co-prescription patterns and associations that provide actionable insights for optimizing medical care and resource allocation.

# Part II: Method Basket Justification

Market Basket Analysis (MBA) is a data mining technique that identifies relationships between items in transactional datasets. In the healthcare context, MBA provides valuable insights into co-prescription patterns among medications, vitamins, and other treatments, helping to reduce readmissions and improve patient outcomes. The Apriori algorithm, a widely used tool for MBA, generates frequent itemsets and association rules using metrics such as support, confidence, and lift to evaluate relationships.(Geeks for Geeks, 2024).

For example, consider a transaction from the dataset representing a patient's prescription history:

* **Transaction ID: 1234**
  + **Prescribed Items:** [Atorvastatin, Glyburide, Amlodipine]

This transaction indicates that the patient was prescribed atorvastatin (a cholesterol-lowering drug), glyburide (a diabetes medication), and amlodipine (used to treat high blood pressure). Using MBA, such transactions are analyzed to uncover co-prescription patterns and associations. For instance, if a high number of transactions include atorvastatin and glyburide together, this could suggest a frequent co-prescription pattern for managing co-morbidities such as diabetes and high cholesterol. This specific example illustrates how transactional data is structured and utilized for identifying actionable insights in the healthcare domain.

By applying MBA, this analysis aims to identify the most frequently co-prescribed medications, uncover patterns that indicate potential synergies or complementary treatments, and derive data-driven recommendations for targeted interventions. To ensure the accuracy of the analysis, this study assumes that each row in the dataset represents a unique patient transaction and that the prescription history is comprehensive. Furthermore, the stability of relationships between prescription items over the analysis period is another critical assumption.

# Part III: Data Preparation and Analysis

The dataset used in this analysis, "medical\_market\_basket.csv," includes 7,501 patient transactions and 20 prescription columns. Data preparation involved several steps to ensure the dataset’s suitability for Market Basket Analysis. Initially, missing values were identified, and rows with all null entries were removed to improve data quality.

1. # Check for any empty values in rows and columns

2. print(f'Are there any empty rows? {data.isnull().values.any()}')

3. print(f'Number of empty columns: {data.columns.isnull().sum()}')

4.

Are there any empty rows? True

Number of empty columns: 0

1. # Drop rows where all elements are NaN

2. data.dropna(how='all', inplace=True)

3. data.shape

4.

(7501, 20)

Next, the dataset was transformed using a Transaction Encoder, which converted it into a binary matrix suitable for MBA. Columns with null transactions were also dropped to maintain data integrity. After cleaning, the final dataset, "cleaned\_data1.csv," consisted of binary indicators representing the presence or absence of each prescription item.

1. from mlxtend.preprocessing import TransactionEncoder

2. # Convert Dataframe to a List of Lists

3. rows = []

4. for i in range (0, data.shape[0]):

5.     rows.append([str(data.values[i,j])

6. for j in range(0, data.shape[1])])

7. # Feed list to Transaction Encoder

8. DE = TransactionEncoder()

9. array = DE.fit(rows).transform(rows)

10.

11. # Return array to DataFrame with column names

12. data1 = pd.DataFrame(array, columns = DE.columns\_)

13. cleaned\_data1 = data1.drop(['nan'], axis=1)

14. cleaned\_data1.head()

15. cleaned\_data1.shape

A screenshot of a computer

Description automatically generated

The Apriori algorithm was applied to this dataset with a minimum support threshold of 0.02 and lift as the primary evaluation metric. Frequent itemsets were generated, highlighting co-occurring prescriptions with significant support, and association rules were extracted to provide insights into prescription patterns. Key metrics, including support, confidence, and lift, were calculated for each rule to assess the strength and significance of the relationships (Data Mining for Business Analytics: Concepts, Techniques, and Applications, 2021).

1. # Import necessary libraries for apriori algorithm

2. from mlxtend.frequent\_patterns import apriori, association\_rules

1. # Create Rules Object

2. rules = apriori(cleaned\_data1, min\_support=0.02, use\_colnames=True)

3. rules.head()

4.

A screenshot of a computer

Description automatically generated

1. # Generate Frequent Itemsets

2. frequent\_itemsets = apriori(cleaned\_data1, min\_support=0.02, use\_colnames=True)

3.

4. # Generate Association Rules

5. rul\_table = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1) # Note: This only works if using mlxtend version 0.23.1 and not the latest 0.23.3

6.

7. # Display Rules

8. rul\_table.head(20)

9.

A table with numbers and letters

Description automatically generated

Figure 1: Rule Table

1. top\_three\_rules = rul\_table.sort\_values(by='lift').head(3)

2. top\_three\_rules

3.

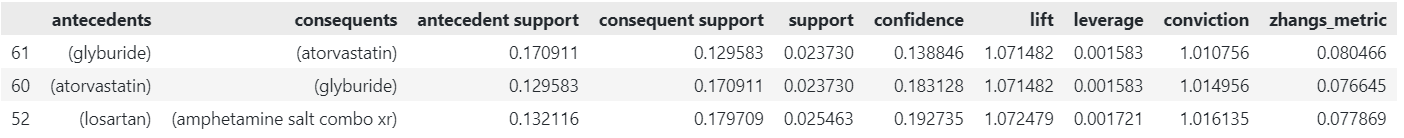


Figure 2: Top Three Rules

The analysis identified three notable rules. The first rule, indicating that patients prescribed glyburide are likely to also be prescribed atorvastatin, had a support value of approximately 2.37%, a confidence level of 13.88%, and a lift of 1.07. This rule suggests a slight positive association, which could reflect specific patient needs. Similarly, the second rule, showing the reverse relationship where atorvastatin precedes glyburide, had the same support value but a higher confidence of 18.31%, further reinforcing this co-prescription pattern. The third rule identified a relationship between losartan and amphetamine salt combo XR, with a support of 2.55%, a confidence of 19.27%, and a lift of 1.07. This rule highlights a potential relationship that warrants further clinical exploration.

# Part IV: Data Summary and Implications

The results of the Market Basket Analysis provide valuable insights into co-prescription patterns among the hospital’s patients. Using the Apriori algorithm, frequent itemsets were identified, and association rules were generated based on support, confidence, and lift metrics. The analysis revealed several key findings:

* **Glyburide and Atorvastatin Association:** Patients prescribed glyburide, a diabetes medication, are frequently co-prescribed atorvastatin, a cholesterol-lowering drug.
  + **Metrics:** Support: 2.37%, Confidence: 13.88%, Lift: 1.07.
  + **Significance:** This association suggests a meaningful relationship that could guide bundled care strategies for diabetic patients managing cardiovascular risks.
* **Reverse Atorvastatin-Glyburide Association:** A higher confidence (18.31%) was observed for atorvastatin preceding glyburide.
  + **Metrics:** Support: 2.37%, Confidence: 18.31%, Lift: 1.07.
  + **Significance:** This bidirectional relationship underscores the importance of tailored treatment plans addressing both diabetes and cholesterol management.
* **Losartan and Amphetamine Salt Combo XR Association:** This pairing, with a support of 2.55% and a confidence of 19.27%, suggests potential overlaps in patient populations managing hypertension and conditions like ADHD.
  + **Metrics:** Support: 2.55%, Confidence: 19.27%, Lift: 1.07.
  + **Significance:** While the lift indicates a weak association, the confidence level highlights the need for further exploration of this pattern in specific patient subgroups.

These findings demonstrate the value of Market Basket Analysis in uncovering co-prescription patterns that support decision-making and improve patient care.

The metrics used in this analysis—support, confidence, and lift—offer critical insights into the relationships:

* **Support:** Measures the frequency of co-prescriptions within the dataset, aiding in operational planning and inventory management.
* **Confidence:** Reflects the likelihood of one prescription leading to another, providing practical insights into prescribing behaviors.
* **Lift:** Evaluates the strength of the association beyond random chance, helping prioritize relationships with potential clinical relevance.

Together, these metrics provide a foundation for evaluating the significance of co-prescription patterns and their practical implications (Data Mining for Business Analytics: Concepts, Techniques, and Applications, 2021).

The practical implications of these findings are significant and include:

* **Optimized Inventory Management:** Frequent co-prescriptions can inform procurement planning, reducing waste and ensuring resources align with patient needs.
* **Bundled Care Strategies:** Associations like glyburide and atorvastatin can guide the development of treatment bundles for managing common co-morbidities.
* **Personalized Treatment Plans:** Relationships such as losartan and amphetamine salt combo XR suggest opportunities for tailored care for specific patient subgroups.

To capitalize on these findings, the following recommendations are proposed:

1. **Integrate Market Basket Analysis:** Use insights from this analysis to inform clinical decisions and improve operational efficiency.
2. **Develop Targeted Care Programs:** Focus on identified patterns to design bundled care approaches, such as pairing glyburide and atorvastatin for diabetic patients.
3. **Conduct Further Analysis:** Explore additional datasets and investigate potential causal relationships to validate findings and uncover more actionable insights.

These recommendations align with the hospital’s strategic goals of improving cost-effectiveness and reducing readmissions through data-driven approaches.

Works Cited

*Data Mining for Business Analytics: Concepts, Techniques, and Applications.* (2021). Hoboken, New Jersey, United States: Wiley. Retrieved December 2024

Geeks for Geeks. (2024, December). Market Basket Analysis in Data Mining. Retrieved from https://www.geeksforgeeks.org/market-basket-analysis-in-data-mining/