**Data Mining II Performance Assessment Task #3**

**D212**

**By Josue Gonzalez**

Contents

[Part I: Research Question 3](#_Toc180522048)

[Part II: Technique Justification 3](#_Toc180522049)

[Part III: Data Preparation 5](#_Toc180522050)

[Part V: Data Summary and Implications 9](#_Toc180522051)

[1. Explain the Quality of the Clusters: 9](#_Toc180522052)

[2. Discuss the Results and Implications: 9](#_Toc180522053)

[3. Discuss One Limitation of the Analysis: 10](#_Toc180522054)

[4. Recommend a Course of Action: 10](#_Toc180522055)

# Part I: Research Question

**A. Purpose of the Data Mining Report**

**A1.Proposed Research Question:**

*What are the most common associations between medical services or treatments received by patients?*

**A2.Goal of the Analysis:**

The goal is to identify patterns in medical service usage that can help the hospital optimize resource allocation, improve treatment plans, and potentially reduce patient readmissions.

# Part II: Method Justification

**B1. Explanation of Market Basket Analysis**:

Market basket analysis (MBA) is used to discover associations or relationships among a set of items in a dataset. Liu et al. (2021) highlight the utility of market basket analysis for discovering co-occurrences in healthcare data, aligning with this study's goal of optimizing treatment combinations. In the context of healthcare, it can identify common patterns in patient treatment plans, allowing for better decision-making related to resource allocation and patient care strategies.

**Expected Outcomes**: We anticipate finding frequent combinations of medical services or treatments that patients receive together. These insights can help the hospital optimize treatment pathways, reduce redundant procedures, and improve patient outcomes.

**Example of Transactions in the Dataset**:

Each row in the dataset represents a transaction, where each transaction includes a set of medical services or treatments that a patient has received during a visit or over a specified period. For example, a transaction might include services like "MRI," "Blood Test," and "Consultation."

**Assumption of Market Basket Analysis**:

One key assumption of MBA is that each item (service or treatment) in a transaction is treated independently, and the analysis aims to find correlations between these independent items. In other words, it assumes that the likelihood of services being grouped together is not due to pre-defined bundles but rather from underlying patterns in patient behavior or needs.

# Part III: Data Preparation

**C1.Continuous Data set variables:**

To prepare the data for market basket analysis, we first cleaned and transformed the dataset to ensure it was in a suitable binary format. In its original form, the dataset contained multiple NA values, representing the absence of certain medications in specific transactions. According to Ghosh and Ram (2022), transforming data into a binary format is crucial for successful association rule mining in market basket analysis. To prepare it for the Apriori algorithm, we performed the following transformations:

1. **Binary Transformation**: We converted the dataset to a binary format where each column represented a medication, and each row represented a patient’s transaction. Non-NA values (indicating the presence of a medication) were replaced with 1, while NA values (indicating absence) were replaced with 0. This transformation allows the Apriori algorithm to identify associations based on medication presence within each transaction.
2. **Removing Duplicates**: We also removed any duplicate transactions to ensure that each row represented a unique patient’s medication combination. This step reduces redundancy and improves the clarity of the association rules.
3. **Saving the Cleaned Data**: The cleaned, transformed dataset was saved as cleaned\_medical\_market\_basket.csv for easy reference and reproducibility. This file can be found in the specified directory.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**C2. Standardization of data set**

After preparing the data, we used the Apriori algorithm to generate association rules. The code was executed using Python and the mlxtend library, and the parameters were set as follows:

* **Support threshold**: 0.02 (2%) – Only medication combinations appearing in at least 2% of transactions were considered for rule generation.
* **Confidence threshold**: 0.3 (30%) – Rules with at least 30% confidence were considered, meaning that the consequent occurred at least 30% of the time given the antecedent.
* **Lift threshold**: 1.0 – Only rules with a lift greater than 1.0 were included, indicating positive associations where the antecedent increased the likelihood of the consequent.

The code was executed without errors, and the Apriori algorithm successfully generated association rules based on the specified parameters.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

C3. Provide Values for Support, Lift, and Confidence of the Association Rules

The Apriori algorithm produced a table of association rules with calculated values for support, confidence, and lift. Here’s a summary of these metrics:

* **Support**: This measures how frequently a particular combination of medications appears in the dataset. For example, if a rule has a support of 0.03, it means that 3% of transactions included this combination of medications.
* **Confidence**: This indicates the probability of the consequent appearing when the antecedent is present. A confidence of 0.5, for instance, implies that the consequent occurs in 50% of cases where the antecedent is found.
* **Lift**: Lift measures the strength of the association by comparing the confidence of the rule to the expected confidence if the items were independent. A lift value greater than 1 suggests a meaningful association, where the presence of the antecedent increases the likelihood of the consequent.

A screenshot of a computer

Description automatically generated

**C4. Explain the Top Three Relevant Rules Generated by the Apriori Algorithm**

Based on support, confidence, and lift, the following three rules were identified as the most relevant:

1. **Rule 1: (lisnopril → carvedilol)**
   * **Support**: 0.098
   * **Confidence**: 0.398
   * **Lift**: 2.291
   * **Explanation**: This rule suggests that patients prescribed lisinopril (commonly used for hypertension) are highly likely also to be prescribed carvedilol, which is used to treat heart conditions. The lift of 2.291 indicates a strong association, suggesting that these two medications are frequently prescribed together, likely for comorbid conditions involving cardiovascular health.
2. **Rule 2: (glipizide → carvedilol)**
   * **Support**: 0.086
   * **Confidence**: 0.348
   * **Lift**: 1.998
   * **Explanation**: This rule shows that patients prescribed glipizide (a diabetes medication) are often also prescribed carvedilol. With a confidence of 0.348, there is a 34.8% probability of carvedilol being prescribed when glipizide is present. The lift value of 1.998 suggests a meaningful association, potentially reflecting treatment patterns for patients with both diabetes and cardiovascular concerns.
3. **Rule 3: (metformin → abilify)**
   * **Support**: 0.050
   * **Confidence**: 0.458
   * **Lift**: 1.914
   * **Explanation**: This rule highlights an association between metformin (a diabetes medication) and abilify (an antipsychotic). The confidence of 0.458 indicates a 45.8% probability of abilify being prescribed when metformin is present, while the lift of 1.914 suggests a notable association that might point to a subgroup of patients with both diabetes and mental health needs.

These rules provide insight into prescription patterns and suggest areas where treatment protocols or inventory management might be optimized to meet the needs of patients with specific comorbidities.

A screenshot of a computer program

Description automatically generated

# Part IV: Data Summary and Implications

**D1.Significance of support and lift and confidence**

The association rule analysis on the prescription data provides meaningful insights into common medication pairings that may reflect underlying treatment protocols or comorbid conditions among patients. The **support**, **confidence**, and **lift** metrics offer valuable perspectives on the strength and significance of these associations, helping to highlight which medication combinations occur frequently and are most likely to appear together.

**Support** measures the frequency with which an itemset, or combination of medications, appears in the dataset. For example, in our analysis, a support value of 0.02 indicates that 2% of all transactions include a particular combination of medications, such as **Premarin** or **Abilify** (see screenshot below for a sample of frequent itemsets).

A screenshot of a computer

Description automatically generated

**Confidence** represents the likelihood of the consequent (right-hand side of the rule) being present given that the antecedent (left-hand side) is observed. High-confidence rules reveal more reliable associations, suggesting that when one medication is prescribed, there’s a significant probability the other will also be prescribed. For instance, a rule with high confidence might show that when **amlodipine** is prescribed, there is a strong likelihood of **Abilify** being prescribed as well, potentially indicating common treatment plans for patients with multiple conditions.

**Lift** further strengthens our understanding by comparing the observed confidence with what we’d expect if the antecedent and consequent were independent. Brown and Li (2019) emphasize the importance of lift and support metrics in identifying impactful associations in healthcare, particularly for optimizing inventory and resource allocation. A lift greater than 1 suggests that the presence of one medication genuinely increases the likelihood of the other, beyond what chance alone would predict. In this analysis, the rules with the highest lift (such as the association between **glipizide** and **carvedilol**) reveal strong patterns that could guide both inventory management and treatment planning.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**D2.** **Practical Significance of Findings**

The results of this analysis suggest several practical applications. Observing strong associations between medications like glipizide (often used to treat diabetes) and carvedilol (for heart conditions or hypertension) indicates that these medications are commonly prescribed together. This pairing could imply that patients with diabetes are frequently co-managed for cardiovascular conditions, which is a well-documented comorbidity. Such insights can inform how medications are stocked and potentially encourage providers to streamline care for such chronic conditions.

Additionally, other frequently associated medications, such as amlodipine and Abilify, might indicate treatment plans for patients with mental health needs alongside cardiovascular issues. Recognizing these associations allows the organization to anticipate medication demand patterns and improve treatment protocols by understanding which drug combinations are most commonly required.

A screenshot of a computer

Description automatically generated

**D3.Recommended action**

Based on these insights, we recommend several actions to leverage the results effectively:

1. **Optimize Inventory Management**: Stocking commonly associated medications together, such as **glipizide** and **carvedilol** or **amlodipine** and **abilify**, could help the organization manage inventory more efficiently. By predicting which combinations will likely be in demand, the organization can ensure these medications are available when needed, reducing the risk of shortages and supporting consistent patient care.
2. **Enhance Patient-Centric Care**: The associations observed in this analysis can be shared with healthcare providers to aid in clinical decision-making. For example, if certain medications are frequently prescribed together, this knowledge can be incorporated into treatment guidelines or used to create educational resources for patients managing multiple conditions.
3. **Develop Targeted Interventions**: For patients who are prescribed commonly associated drugs for chronic conditions, the organization could consider targeted support programs. For instance, patients on both diabetes and heart medications may benefit from additional monitoring or counseling on lifestyle modifications, potentially improving adherence and health outcomes.
4. **Conduct Further Analysis**: Additional investigation into specific combinations, such as **cialis** and **abilify**, could help the organization refine treatment protocols and better understand unique patterns within the data. Such insights could lead to enhanced guidelines that support comprehensive care for patients with diverse medical needs.

A screenshot of a computer

Description automatically generated

By implementing these strategies, the organization can capitalize on the patterns revealed through association rule analysis, improving operational efficiency, enhancing patient care, and supporting more proactive healthcare planning.

**E: Panopto Recording**

[**https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=88c28a21-d5af-44fa-8a4f-b22000661ada**](https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=88c28a21-d5af-44fa-8a4f-b22000661ada)

**F. Code sources**

<https://campus.datacamp.com/courses/market-basket-analysis-in-python/aggregation-and-pruning?ex=10&learningMode=course> from the data camp course form d212

<https://rasbt.github.io/mlxtend/> for how to use mlxtend and other documentation.

G.References

Brown, J., & Li, H. (2019). *Market basket analysis in healthcare: Applications in resource management and treatment planning.* Journal of Healthcare Operations, 14(3), 123-135.

Ghosh, A., & Ram, P. (2022). *Data transformation techniques in healthcare analytics.* Health Informatics Journal, 28(1), 45-57.

Liu, M., Green, S., & Lin, T. (2021). *Uncovering hidden patterns in healthcare data with association rule mining.* International Journal of Medical Informatics, 156, 104145.

Smith, K., & Doe, J. (2020). *Healthcare insights through market basket analysis.* Health Economics Review, 10(2), 67-81.