D212 – Data Mining II

WGU M.S. Data Analytics

Lyssa Kline

March 11, 2025

A, Part 1: Research Question

A1, Research Question and Technique

The research question in scope for this analysis is " What are the most frequently prescribed medication combinations among readmitted patients, and how do these prescriptions relate to specific medical conditions? This question helps to identify patterns in prescription data, revealing which medications tend to be administered together for patients who experience hospital readmissions. By understanding these associations, the hospital can improve medication management to reduce readmissions, identify risky medication combinations that may lead to complications, and optimize treatment protocols for common conditions leading to readmissions.

A2, Analysis Goal

The primary goal of this analysis is to use a Market Basket analysis to identify associations between prescribed medications and readmitted patients to optimize treatment plans and reduce readmission rates. This goal aligns with hospitals' long-term cost-effectiveness strategy as reducing readmissions leads to more efficient resource allocation and better patient outcomes. By analyzing hospital prescription data, the hospital can develop data-driven interventions such as flagging high-risk medications, adjusting treatment plans, and providing personalized patient care.

B, Part 2: Market Basket Justification

B1, Market Basket Technique and Outcomes

Market Basket Analysis (MBA) is a data mining technique to identify associations between items in transactional datasets. In this case, it will be used to analyze prescription data and uncover relationships between commonly prescribed medications for re-admitted patients.

The dataset includes prescriptions (Presc01 and Presc20) given to patients, each row represents a patient's prescription history or transaction. Market basket analysis will identify patterns in co-occurring prescriptions, showing which medications are frequently prescribed together.

Expected outcomes are to identify frequently prescribed medication combinations, detect prescription patterns for readmitted patients, and optimize medication protocols. In observing these outcomes, hospitals be able to start to understand which medications are associated with higher-risk conditions and review ways hospitals can reduce readmission rates by altering prescription patterns.

B2, Transaction Example

Each row represents a patient's medical record, where each column indicates whether the patient received that medication.
Here's an example transaction where the patient has the first 5 prescriptions in the dataset:

```
     Presc01             Presc02      Presc03       Presc04    Presc05  \
1  amlodipine  albuterol aerosol  allopurinol  pantoprazole  lorazepam
```

This relates to MBA because this patient has diabetes, high blood pressure, stroke, and hyperlipidemia. MBA will check how often these conditions appear together across many patients. This may lead to finding a strong association that leads to a risk of stroke or frequent emergency visits.

B3, Assumption

One key assumption of market-basket analysis is the independence of transactions. Each transaction is independent of the others. MBA assumes that each patient's prescription set is independent, meaning that one patient's prescriptions do not directly influence another patient. However, patients with similar medical conditions may receive similar treatments, making some prescriptions dependent on medical guidelines. Despite this, MBA still helps identify important relationships in prescribing patterns, advising better decision-making for education management.

Market Basket Analysis is an effective way to analyze prescription patterns, identify risk factors for readmissions, and optimize medication usage in hospitals.

C, Part 3: Data Preparation

C1, Cleaned Dataset

A prepared dataset outputted in CSV format can be found within the attached 'prepared_medical_task3.csv' file.

C2, Apriori Algorithm Code

The apriori algorithm finds frequent itemset and generates association rules, the expected output is that the association rules table will be displayed, showing support, confidence, and lift for each rule. The screenshot below shows the code used to generate the association rules with the error-free execution of the apriori algorithm. As you can see below, the code correctly executed Apriori to generate association rules between medications.

```
# Ensure prescription columns are boolean (True/False)
df_cleaned[presc_columns] = df_cleaned[presc_columns].astype(bool)

# Run Apriori Algorithm to find frequent itemsets
frequent_itemsets = apriori(df_cleaned[presc_columns], min_support=0.01, use_colnames=True)

# Generate association rules
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)

# Display the first few rows of association rules
print(rules.head())
```

```
  antecedents consequents  antecedent support  consequent support   support  \
0   (Presc01)   (Presc02)            0.500000            0.383082  0.383082
1   (Presc02)   (Presc01)            0.383082            0.500000  0.383082
2   (Presc01)   (Presc03)            0.500000            0.292561  0.292561
3   (Presc03)   (Presc01)            0.292561            0.500000  0.292561
4   (Presc01)   (Presc04)            0.500000            0.222970  0.222970

   confidence  lift  representativity  leverage  conviction  zhangs_metric  \
0    0.766165   2.0               1.0  0.191541    2.638255       1.000000
1    1.000000   2.0               1.0  0.191541         inf       0.810481
2    0.585122   2.0               1.0  0.146280    1.705174       1.000000
3    1.000000   2.0               1.0  0.146280         inf       0.706775
4    0.445941   2.0               1.0  0.111485    1.402430       1.000000

    jaccard  certainty  kulczynski
0  0.766165   0.620962    0.883082
1  0.766165   1.000000    0.883082
2  0.585122   0.413549    0.792561
3  0.585122   1.000000    0.792561
4  0.445941   0.286952    0.722970
```

C3, Association Rules

After executing the Apriori algorithm, we can start to review the values for support, lift, and confidence. The screenshot below shows the output of the association rules tables showing the values. Support measures how frequently a prescription combination appears in the dataset. Confidence indicates how likely a consequent medication is prescribed when an antecedent medication is given. Lift measures the strength of an association compared to random chance.

```
# Select relevant columns - providing value for support, lift, and confidence
rules_metrics = rules[['antecedents', 'consequents', 'support', 'confidence', 'lift']]

print(rules_metrics)
```

```
       antecedents                                         consequents  \
0        (Presc01)                                           (Presc02)
1        (Presc02)                                           (Presc01)
2        (Presc01)                                           (Presc03)
3        (Presc03)                                           (Presc01)
4        (Presc01)                                           (Presc04)
...            ...                                                 ...
523245   (Presc02)  (Presc09, Presc03, Presc12, Presc08, Presc01, ...
523246   (Presc11)  (Presc09, Presc03, Presc12, Presc08, Presc01, ...
523247   (Presc07)  (Presc09, Presc03, Presc12, Presc08, Presc01, ...
523248   (Presc10)  (Presc09, Presc03, Presc12, Presc08, Presc01, ...
523249   (Presc04)  (Presc09, Presc03, Presc12, Presc08, Presc01, ...

         support  confidence       lift
0       0.383082    0.766165   2.000000
1       0.383082    1.000000   2.000000
2       0.292561    0.585122   2.000000
3       0.292561    1.000000   2.000000
4       0.222970    0.445941   2.000000
...          ...         ...        ...
523245  0.010265    0.026797   2.610405
523246  0.010265    0.601563  58.601563
523247  0.010265    0.112491  10.958364
523248  0.010265    0.389873  37.979747
523249  0.010265    0.046039   4.484903

[523250 rows x 5 columns]
```

In this output for support, if presc01 and presc02 appear in 38.3% of transactions, then support = 0.383082. In this output for confidence, a confidence of 1.0 means whenever presc12 is prescribed, presc11 is also prescribed. Lastly, for lift, a lift of 58.60 means that presc11 is 58.6 times more likely to be prescribed when presc12 is prescribed, compared to random prescribing.

The association rules correctly include supporting confidence and lifting and confirming strong medication prescription patterns.

C4, Relevant Rules

When analyzing support, life, and confidence we can start to review the top three strongest association rules, sorted by lift and confidence. A discussion of the top three rules can be found below the screenshot.

```
# Sort rules by Lift and Confidence - explain top 3 relevant rules
top_rules = rules.sort_values(by=['lift', 'confidence'], ascending=False).head(3)

# Display top 3 rules
print(top_rules)
```

```
            antecedents          consequents  antecedent support  \
130           (Presc12)            (Presc11)            0.010265
456  (Presc01, Presc12)            (Presc11)            0.010265
460           (Presc12)   (Presc01, Presc11)            0.010265

     consequent support   support  confidence       lift  representativity  \
130            0.017064  0.010265         1.0  58.601563               1.0
456            0.017064  0.010265         1.0  58.601563               1.0
460            0.017064  0.010265         1.0  58.601563               1.0

     leverage  conviction  zhangs_metric   jaccard  certainty  kulczynski
130   0.01009         inf        0.99313  0.601563        1.0    0.800781
456   0.01009         inf        0.99313  0.601563        1.0    0.800781
460   0.01009         inf        0.99313  0.601563        1.0    0.800781
```

Rule 1 provides findings that Presc11 is highly likely when Presc12 is given. The support shows that 1.02% of all transactions involve this combination. The confidence level is 1.0, meaning that 100% of the time presc12 is given, presc11 is also given. The lift is 58.60 meaning that presc11 is 58.6 times more likely to be prescribed if presc12 is also given. From these stats, we can conclude that this rule shows that prescription 12 and prescription 11 are strongly linked medications and are often prescribed together.

Rule 2 provides findings that if Presc01 and Presc12 are given, Presc11 is always prescribed. The support shows that 1.02% of all transactions involve this combination. The confidence level is 1.0, meaning that 100% of the time when presc01 and presc12 are prescribed, presc11 is also given. The lift is 58.60 meaning that this rule is extremely strong. From these stats, we can conclude that when a patient is given prescription 1 and prescription 12, then prescription 11 is almost always required.

Rule 3 provides findings that if Presc12 is given, then Presc01 and Presc11 are also given. The support shows that 1.02% of all transactions involve this combination. The confidence level is 1.0, meaning that 100% of the time when presc12 is prescribed, presc01 and presc11 are also prescribed. The lift is 58.60 meaning that this rule is extremely strong. From these stats, we can conclude that prescription 12 is a key indicator that a patient will also receive prescription 1 and prescription 11.

Overall, prescription 12 (Presc12) is the most influential medication because it is present in all three rules. Prescription 11 (Presc11) is strongly associated with prescription 12 (Presc12) meaning it is almost always prescribed together. The hospital can use this data to optimize medication stocking and patient treatment plans.

D, Part 4: Data Summary and Implications

D1, Significance Summary, and Results

Support, confidence, and lift are key measures in Market Basket Analysis that help evaluate the strength of relationships between medications.

Support measures how often a particular medication combination appears in the dataset. For example, from this analysis, we saw that prescription 12 and prescription 11 have a support of 1.02% meaning that 1.02% of all patients received both prescription 12 and prescription 11. Higher support suggests a frequently prescribed combination and lower support may indicate rare but potentially important prescriptions.

Confidence measures how often the consequent medication is prescribed when the antecedent medication is prescribed. For example, the prescription 12 and prescription 11 combination have a confidence of 100%. This means that every time prescription 12 is prescribed, prescription 11 is also prescribed. A high confidence suggests a strong prescribing pattern, this information could be used by hospitals to predict prescriptions and plan inventory.

Lift measures how much more likely two medications are prescribed together compared to if they were given randomly. In our prescription 12 and prescription 11 group example, this has a lift of 58.60. This means that prescription 11 is 58.6 times more likely to be prescribed when prescription 12 is given. A lift > 1 indicates a strong positive relationship. Higher lift values suggest a critical medication pair, meaning doctors frequently prescribe them together.

The key takeaway from this summary and findings is that prescription 12 and prescription 11 have a very strong association, meaning they are almost always prescribed together. High support, confidence, and lift validate important prescription patterns.

D2, Practical Significance and Findings

Our findings reveal important insights for hospital operations and patient care. This consists of medication management and stocking, prescription pattern optimization, risk of over-prescription or drug interactions, and cost and insurance optimization.

The knowledge of prescription 12 and prescription 11 being highly correlated, could allow hospitals to use this information and ensure that these medications are stocked together to avoid shortages and improve treatment efficiency. Since prescription 12 always leads to prescription 11, hospitals could automate medication ordering systems to suggest prescription 11 and prescription 12 is prescribed. this could help train doctors on prescribing trends to improve efficiency. Additionally, when reviewing this information and finding if certain medications are always prescribed together, hospitals can start to evaluate whether these combinations are medically necessary. This could help draw significance to the need for further research on the potential side effects of high-confidence prescription pairs. Lastly, if specific medications are always given together, this could allow hospitals to negotiate better pricing from suppliers. This could allow for significant cost-cutting measures, allowing insurance companies to develop better coverage plans for common treatment patterns.

This analysis gives a significant overview of ways that hospitals can begin to optimize stocking, improve prescription workflows, and reduce unnecessary medication costs.

D3, Course of Action

Based on the findings of the market basket analysis performed on prescription drugs, the hospital should take the following three strategic actions.

First, implementing a predictive prescription system. Hospitals could implement the use of AI or rule-based alerts to automatically suggest medications based on strong associations. For example, if a doctor prescribes prescription 12, the system will automatically suggest adding prescription 11. This would allow doctors to save time, reduce prescription errors, and improve patient care consistency.

Second, the hospital should optimize medication inventory based on market basket analysis. For example, since prescription 12 and prescription 11 are always prescribed together hospitals can stock these medications together in pharmacies. Using forecasting models to predict demand, would ensure availability to prevent delays in treatment.

Lastly, hospitals should investigate high-lift medication combinations for safety. Conducting further clinical research on whether highly associated medications (e.g. prescription 12 and prescription 11) should always be prescribed together would allow hospitals to identify potential risks of overprescription and improve patient safety guidelines.

In implementing these action items, hospitals can start to optimize prescription plans to ensure the patient is getting the best care with the least delays in treatment.

Part 5: Attachments

E, Panopto Video Recording

The video recording for this assignment includes a vocalized demonstration of all the code, the code being executed, and the results of the code being represented inside this report.
The video recording for this project can be found inside the Panopto drop box titled “D212-Kline-Task3”

Panopto video link: https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=36645127-602f-44bd-845b-b29d0172e5c7

F, Web Sources

No sources or segments of third-party code were used to acquire data or to support the report.

G, Acknowledge the Sources

I acknowledge that no segments of third-party sources were directly stated or copied from the web into this report.