Rancel Hernandez

D212 Data Mining

Jan 28, 2025

ASSOCIATION RULES AND LIFT ANALYSIS

A1: PROPOSAL OF QUESTION

What are the most significant and strongest associations between medication classifications prescribed to patients, and how frequently do these combinations occur?

A2: DEFINED GOAL

A goal of this analysis is to explain the correlation between the most frequent and strongest associations between medication classifications. For example, if the likelihood of Classification A being prescribed increases when Classification B is prescribed, then it may indicate a meaningful relationship. Identifying the most common classifications can help to efficiently explore prescription patterns and assess the effectiveness of these combinations in relation to readmission rates. Additionally, understanding how common prescription practices relate to readmissions can guide doctors in prescribing combinations within the same class that are correlated with lower readmission rates. This goal is both relevant to the analysis and within the scope of the available data.

B1: EXPLANATION OF MARKET BASKET

A market basket analysis aims to identify meaningful associations between item sets and relies on transaction datasets, where each entry represents a collection of items. Due to the large number of potential combinations, the analysis typically prunes less frequent sets to focus on the most common ones. This is often achieved using the Apriori algorithm that calculates the frequency of individual items, known as support, and removes item sets containing these uncommon items because they are less likely to form meaningful associations [1].

Once the most frequent item sets are identified, they are used to create association rules, with one set acting as the predictor, known as the antecedent, and the other as the outcome, known as the consequent. These rules indicate how the presence of one set influences the likelihood of another, either positively or negatively. Metrics such as support, confidence, and lift are then calculated to assess the strength of these associations. Confidence measures the likelihood of the consequent given the antecedent, while lift compares the likelihood of the consequent with and without the antecedent to provide an overall strength of the association [2].

In this analysis, the dataset consists of combinations of medications prescribed to patients, and each patient can be prescribed up to 20 medications from a list of 119 unique options. To reduce the dimensionality from 119 down to 30, the medications were aggregated based on their common uses using various medication classification sources such as the Federal

Drug Administration (FDA), National Library of Medicine, and the Norwegian Institute of Public Health [3][4][5][6]. These classifications are shown in *Table 1*.

This transformation preserves prescription patterns by focusing on the general use of similar drugs rather than on individual medications. The dataset was restructured so that columns represent medication classifications, with a true value indicating the presence of a classification in a patient's transaction and false otherwise. The most frequent item sets were identified to create association rules, which were then filtered and evaluated based on support, confidence, and lift to determine the top three optimal rules consisting of classification sets.

Table 1: Medication Classifications

Classification	Description	Medications
Analgesics	Pain-relieving medications	hydrocodone, oxycodone, codeine, tramadol, acetaminophen
Antianxiety Drugs	Manage anxiety and related conditions	alprazolam, clonazepam, diazepam, lorazepam, temezepam, zolpidem
Antiarrhythmics	Manage abnormal heart rhythms	verapamil SR
Antibiotics	Treat bacterial infections	amoxicillin, cefdinir, cephalexin, ciprofloxacin, levofloxacin, clavulanate K+, doxycycline hyclate, azithromycin, sulfamethoxazole, trimethoprim DS
Anticoagulants	Prevent blood clot formation	clopidogrel
Anticonvulsants	Manage neurological conditions	gabapentin, pregabalin, topiramate
Antidepressants	Manage depression and anxiety disorders	citalopram, escitalopram, fluoxetine HCI, paroxetine, sertraline HCI, duloxetine, venlafaxine XR, cymbalta, amitriptyline, bupropion SR, trazodone HCI

Antifungals	Treat fungal infections	clotrimazole, fluconozole
Antihistamines	Alleviate symptoms of allergies or colds	fexofenadine, promethazine
Antihypertensives	Control high blood pressure	benazepril, enalapril, lisinopril, atenolol, carvedilol, metoprolol, metoprolol succinate XL, metoprolol tartrate, amlodipine, benicar, losartan, valsartan, clonidine HCI
Anti-Inflammatories	Reduce inflammation	celebrex, celecoxib, diclofenac sodium, ibuprophen, meloxicam, naproxen, methylprednisone, prednisone, triamcinolone Ace topical, hydrocortisone 2.5% cream
Antipsychotics	Manage severe mental health disorders	abilify, quetiapine
Antivirals	Treat viral infections	valaciclovir
Bronchodilators	Respiratory condition management	albuterol HFA, albuterol aerosol, salmeterol inhaler
Corticosteroids	Reduce inflammation, especially for respiratory conditions	flovent HFA 110 mcg inhaler, fluticasone, fluticasone nasal spray, mometasone
Diuretics	Promote fluid loss to manage blood pressure	furosemide, hydrochlorothiazide, spironolactone, triamterene
Hormones	Hormonal treatments	levothyroxine sodium, synthroid, Premarin, Yaz
Hypoglycemics	Manage blood sugar levels in diabetic patients	metformin, metformin HCI, glimepiride, glipizide, glyburide, pioglitazone, lantus
Muscle Relaxants	Relax skeletal muscles to reduce spasms	carisoprodol, cyclobenzaprine
Vitamins	Essential nutrients	folic acid
Antacids	Reduce stomach acid	ranitidine
Antilipidemics	Lower cholesterol and triglyceride levels	atorvastatin, crestor, ezetimibe, fenofibrate, lovastatin, pravastatin, rosuvastatin, simvastatin
Antianginals	Relieve chest pain	isosorbide mononitrate
Bone Health	Help maintain bone density	actonel, alendronate, boniva
Electrolytes	Maintain electrolyte	potassium chloride

	balance	
Erectile Dysfunction	Treat erectile dysfunction	cialis, viagra
Gout Medications	Reduce uric acid levels to manage gout	allopurinol
Proton Pump Inhibitors	Reduce stomach acid by blocking proton pumps	esomeprazole, lansoprazole, omeprazole, pantoprazole
Prostate	Manage symptoms related to prostate enlargement	finasteride, tamsulosin
Stimulants	Increase focus or alertness for conditions like ADHD	amphetamine, amphetamine salt combo, amphetamine salt combo XR, dextroamphetamine XR

Medication classifications were sourced from the FDA [3, 4], the National Library of Medicine [5], and the Norwegian Institute of Public Health [6].

B2: TRANSACTION EXAMPLE

The third entry of the dataset is provided as an example both before and after the data transformation. This transaction includes three medications: citalopram, benicar, and amphetamine salt combo XR. The remaining columns contain NaN values that indicate no additional medications were prescribed. After the transformation, these medications are aggregated into their corresponding classifications: citalopram under antidepressants, benicar under antihypertensives, and amphetamine salt combo XR under stimulants. The columns represent the medication limit for each transaction before the transformation, and they represent medication classifications after the transformation. This process is illustrated in *Figures 1 and 2*.

```
Presc01                      citalopram
Presc02                         benicar
Presc03    amphetamine salt combo xr
Presc04                             NaN
Presc05                             NaN
Presc06                             NaN
Presc07                             NaN
Presc08                             NaN
Presc09                             NaN
Presc10                             NaN
Presc11                             NaN
Presc12                             NaN
Presc13                             NaN
Presc14                             NaN
Presc15                             NaN
Presc16                             NaN
Presc17                             NaN
Presc18                             NaN
Presc19                             NaN
Presc20                             NaN
Name: 3, dtype: object
```

Figure 1: A depiction of the third transaction in the dataset before any transformation.

```
Analgesics                  False
Antianxiety Drugs           False
Antiarrhythmics             False
Antibiotics                 False
Anticoagulants              False
Anticonvulsants             False
Antidepressants              True
Antifungals                 False
Antihistamines              False
Antihypertensives            True
Anti-Inflammatories         False
Antipsychotics              False
Antivirals                  False
Bronchodilators             False
Corticosteroids             False
Diuretics                   False
Hormones                    False
Hypoglycemics               False
Muscle Relaxants            False
Vitamins                    False
Antacids                    False
Antilipidemics              False
Antianginals                False
Bone Health                 False
Electrolytes                False
Erectile Dysfunction        False
Gout Medications            False
Proton Pump Inhibitors      False
Prostate                    False
Stimulants                   True
Name: 3, dtype: bool
```

Figure 2: A depiction of the third transaction in the dataset after the aggregation.

B3: MARKET BASKET ASSUMPTION

An assumption of a market basket analysis is that the transactions must contain at least two items because the goal is to identify meaningful correlations between item sets, known as association rules. These rules consist of an antecedent and a consequent, so if all transactions contain only one item, no meaningful associations can be made. Additionally, the metrics required for the analysis, such as support, cannot be calculated because support depends on identifying associations across multiple items. Therefore, for association rules to be generated and meaningful insights to be derived, each transaction must contain at least two items.

C1: TRANSFORMING THE DATA SET

The cleaned dataset will be submitted in a CSV file.

C2: CODE EXECUTION

The creation of the item sets using the Apriori algorithm are shown in *Figure 3*. A minimum support of 0.001 was set as the threshold to exclude less frequent item sets. The association rules generated from these item sets are displayed in *Figure 4*, and a minimum support of 0 was initially used to plot the rules and determine the optimal threshold.

```
In [14]: frequent_items = apriori(onehot_classes, use_colnames=True, min_support=0.001,  max_len=4)
```

```
In [15]: frequent_items.head()
```

Out[15]:

	support	itemsets
0	0.034129	(Analgesics)
1	0.131049	(Antianxiety Drugs)
2	0.002800	(Antiarrhythmics)
3	0.101520	(Antibiotics)
4	0.029996	(Anticoagulants)

Figure 3: A depiction of the creation of the itemsets using the Apriori algorithm.

```
In [16]: rules = association_rules(frequent_items, metric='support', min_threshold=0.0)
```

```
In [17]: rules.head()
```

Out[17]:

consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage	conviction	zhangs_metric	jaccard	certainty	kulczynski
(Antianxiety Drugs)	0.034129	0.131049	0.012598	0.369141	2.816810	1.0	0.008126	1.377408	0.667779	0.082569	0.273999	0.232637
(Analgesics)	0.131049	0.034129	0.012598	0.096134	2.816810	1.0	0.008126	1.068600	0.742261	0.082569	0.064196	0.232637
(Antibiotics)	0.034129	0.101520	0.007266	0.212891	2.097036	1.0	0.003801	1.141493	0.541621	0.056594	0.123955	0.142230
(Analgesics)	0.101520	0.034129	0.007266	0.071569	2.097036	1.0	0.003801	1.040327	0.582246	0.056594	0.038763	0.142230
(Anticoagulants)	0.034129	0.029996	0.002466	0.072266	2.409175	1.0	0.001443	1.045562	0.605588	0.040000	0.043577	0.077244

```
In [18]: rules.shape
Out[18]: (22354, 14)
```

Figure 4: A depiction of the association rules created from the item sets, and 22,354 rules were created with a minimum threshold of 0.

C3: ASSOCIATION RULES TABLE

The association rules table is shown in *Figure 5*, and it displays the support, lift, and confidence of the rules generated using the optimal minimum support threshold of 0.05.

	antecedents	consequents	support	lift	confidence
0	(Antihypertensives)	(Antianxiety Drugs)	0.069591	2.242182	0.293836
1	(Antianxiety Drugs)	(Antihypertensives)	0.069591	2.242182	0.531027
2	(Antibiotics)	(Antihypertensives)	0.061125	2.542280	0.602101
3	(Antihypertensives)	(Antibiotics)	0.061125	2.542280	0.258092
4	(Antidepressants)	(Antihypertensives)	0.062658	2.494662	0.590823
5	(Antihypertensives)	(Antidepressants)	0.062658	2.494662	0.264565
6	(Anti-Inflammatories)	(Antihypertensives)	0.071657	2.659065	0.629760
7	(Antihypertensives)	(Anti-Inflammatories)	0.071657	2.659065	0.302561
8	(Antihypertensives)	(Antipsychotics)	0.074990	2.607102	0.316634
9	(Antipsychotics)	(Antihypertensives)	0.074990	2.607102	0.617453
10	(Antihypertensives)	(Hypoglycemics)	0.084122	2.323856	0.355193
11	(Hypoglycemics)	(Antihypertensives)	0.084122	2.323856	0.550371
12	(Antilipidemics)	(Antihypertensives)	0.077923	2.308664	0.546773
13	(Antihypertensives)	(Antilipidemics)	0.077923	2.308664	0.329018
14	(Antihypertensives)	(Stimulants)	0.085655	2.321659	0.361666
15	(Stimulants)	(Antihypertensives)	0.085655	2.321659	0.549850
16	(Antilipidemics)	(Hypoglycemics)	0.051860	2.380768	0.363891
17	(Hypoglycemics)	(Antilipidemics)	0.051860	2.380768	0.339294
18	(Hypoglycemics)	(Stimulants)	0.056726	2.382407	0.371130
19	(Stimulants)	(Hypoglycemics)	0.056726	2.382407	0.364142

Figure 5: A depiction of the association rules table that includes the support, lift, and confidence.

C4: TOP THREE RULES

The top three rules generated from the Apriori algorithm are as follows: (Antibiotics) => (Antihypertensives), (Anti-Inflammatories) => (Antihypertensives), and (Antipsychotics) => (Antihypertensives). These rules were determined to be at the top by filtering the association rules table by the support, lift, and confidence metrics. The thresholds used were 0.06 for support, 2.5 for lift, and 0.5 for confidence. High threshold values were selected to ensure the identification of the most significant and reliable rules. The analysis prioritized lift and confidence to assess the strength of the associations, and then support to evaluate their frequency. These top three rules can be seen in *Figure 6*.

The first rule indicates that approximately 60% of patients who were prescribed antibiotics also received antihypertensives, and this percentage is derived from the confidence value of 0.602101. Furthermore, the lift value of around 2.5 suggests that patients prescribed antibiotics were also 2.5 times more likely to be prescribed antihypertensives than by random chance [2]. This association is fairly common because it appears in approximately 6% of all transactions in the dataset, and this is indicated by the support value of 0.061125. Therefore, the high confidence and lift values are evidence of a strong and frequent association.

Antibiotics are medications used to treat bacterial infections, and antihypertensives are used to manage high blood pressure. A possible explanation for this association is that individuals with hypertension often experience chronic inflammation that can weaken their immune system over time [7]. A weakened immune system may increase susceptibility to bacterial infections, leading to a higher likelihood of antibiotic prescriptions [8]. Therefore, the frequent combination of antibiotics and antihypertensives could reflect this relationship between hypertension, immune function, and infection risk.

The second rule between anti-inflammatories and antihypertensives had a confidence value of approximately 0.63, indicating that 63% of patients who were prescribed anti-inflammatories also received antihypertensives. This is further supported by the high lift value of around 2.7, suggesting that patients prescribed anti-inflammatories were about 2.7 times more likely to receive antihypertensives than by random chance. Together, these metrics demonstrate a strong association between these two classifications. Additionally, the support value of approximately 0.072 indicates that this association was fairly common and was present in 7.2% of the transactions in the dataset. Anti-inflammatories are medications used to reduce inflammation, swelling, and pain. As mentioned earlier, patients prescribed antihypertensives typically suffer from hypertension or other cardiovascular conditions that can lead to chronic inflammation due to the increased blood pressure [7]. Thus, the strong association between these classifications is expected.

The third rule between antipsychotics and antihypertensives has a confidence value of approximately 0.62, meaning that 62% of patients prescribed antipsychotics also received antihypertensives. This high confidence is further supported by the lift value of around 2.61, which indicates that patients prescribed antipsychotics were 2.61 times more likely to be prescribed antihypertensives than by random chance. Additionally, the association between these two classifications is the most frequent among the top three rules because it had the highest support value of approximately 0.075, indicating that this combination appeared in 7.5% of the transactions.

Antipsychotics are used to treat mental disorders that affect a person's mood, thoughts, or behavior [9]. In this classification, there is a medication Quetiapine, also known as Seroquel, that has been linked to an increase in high blood pressure in children and adolescents [10]. Therefore, the increased blood pressure side effect of antipsychotics may explain the association between antipsychotics and antihypertensives in the dataset.

The top three rules were (Antibiotics) => (Antihypertensives), (Anti-Inflammatories) => (Antihypertensives), and (Antipsychotics) => (Antihypertensives) because their strong associations were supported by their high lift and confidence values. Additionally, these rules were frequent in the dataset, with high support values reflecting their relevance and plausible explanations further supporting their relationship.

```
            antecedents        consequents  antecedent support  \
2           Antibiotics  Antihypertensives            0.101520
6   Anti-Inflammatories  Antihypertensives            0.113785
9        Antipsychotics  Antihypertensives            0.121450

   consequent support   support      lift  confidence
2            0.236835  0.061125  2.542280    0.602101
6            0.236835  0.071657  2.659065    0.629760
9            0.236835  0.074990  2.607102    0.617453
```

Figure 6: A depiction of the top three association rules.

D1: SIGNIFICANCE OF SUPPORT, LIFT, AND CONFIDENCE SUMMARY

The analysis resulted in generating association rules with high support, lift, and confidence values. The support metric reflects the frequency of the rules in the transactions, and the average support of all the rules was approximately 0.07, meaning that these rules were all common and appeared in around 7% of the transactions. In this context, the support is high because there are 30 different classifications and a large number of possible combinations, but the ones identified were present in nearly 1 in 10 transactions.

Since these rules were established as frequent, assessing the strength of the combinations revealed their validity. The confidence indicated the percent of transactions that given the consequent the antecedent was also present, and the average confidence of the rules was approximately 42%. While this is on the lower end, the lower confidence of some rules may stem from the consequent having a higher frequency of 0.24 compared to the antecedent's support of 0.1. This suggests that, due to the consequent being frequent and associated with other classifications, the overall combinations were less likely to appear together distinctly. However, the minimum lift value of around 2.2 validated and ensured that all rules had meaningful relationships and were not due to random chance, despite the lower confidence [2]. As a result, the analysis was able to identify strong, meaningful, and frequent associations in the dataset.

D2: PRACTICAL SIGNIFICANCE OF FINDINGS

The practical significance of the findings from the analysis identified valid and frequent associations between medication classifications in the dataset. These associations reflect common prescription practices of doctors in the network. By focusing on the most frequent combinations of classifications, a targeted analysis can explore relationships for these common medication types. For example, a market basket analysis of medications in the third top rule, (Antipsychotics) => (Antihypertensives), revealed various combinations involving Abilify. Evaluating these combinations can help identify those correlated with lower readmissions that are more effective treatment options. The insights have broader implications for healthcare optimization, potentially informing more personalized treatment strategies, refined prescription guidelines, and targeted intervention protocols. Ultimately, this analysis can guide improved prescription practices, enhance patient outcomes, and reduce readmissions.

D3: COURSE OF ACTION

Based on the results of the analysis, it is recommended that the network further explore the relationships between medications from classifications that are strongly associated and frequent. Their strong associations justify meaningful correlations, and focusing on the most frequent ones allows for a targeted and efficient assessment of these combinations. They can be assessed by comparing readmission rates between medication combinations as a measure of effectiveness and exploring how they differ across demographics such as age and gender. Additionally, these targeted analyses focus on associations between medications within a select subset of classes, with combinations that may have been overlooked when considering all medications together. A good starting point could be to prioritize the top three rules supported by high support, lift, and confidence values because they can help to efficiently reduce readmissions, improve prescription practices, and enhance patient outcomes. These insights are valuable because the Centers for Medicare and Medicaid Services penalize hospitals for excessive readmissions. Therefore, by further exploring these associations, the network can also avoid fines while improving overall patient care.

F: SOURCES FOR THIRD-PARTY CODE

NumPy, "numpy.logical_or," NumPy v2.1 Documentation. [Online]. Available: https://numpy.org/doc/2.1/reference/generated/numpy.logical_or.html. [Accessed: Jan. 30, 2025].

M. L. Waskom, "seaborn.heatmap," Seaborn Documentation. [Online]. Available: https://seaborn.pydata.org/generated/seaborn.heatmap.html. [Accessed: Jan. 30, 2025].

G: SOURCES

Reference List

[1] S. Raschka, "mlxtend: Apriori," mlxtend Documentation. [Online]. Available: https://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/. [Accessed: Jan. 28, 2025].

[2] S. Vijayaraghavan, "Association Rules 2: The Apriori Algorithm," Towards Data Science, 2021. [Online]. Available: https://towardsdatascience.com/association-rules-2-aa9a77241654. [Accessed: Jan. 28, 2025].

[3] U.S. Food and Drug Administration, "General Drug Categories," FDA. [Online]. Available: https://www.fda.gov/drugs/investigational-new-drug-ind-application/general-drug-categories. [Accessed: Jan. 29, 2025].

[4] ATC/DDD Index, “ATC/DDD Index,” *Norwegian Institute of Public Health*. [Online]. Available: https://atcddd.fhi.no/atc_ddd_index/. [Accessed: Jan. 30, 2025].

[5] U.S. National Library of Medicine, "Drug Information: Drug Classes," MedlinePlus. [Online]. Available: https://medlineplus.gov/druginformation.html. [Accessed: Jan. 29, 2025].

[6] U.S. Food and Drug Administration (FDA), "Drug Approval Packages," FDA. [Online]. Available: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm. [Accessed: Jan. 29, 2025].

[7] Z. Zhang, L. Zhao, X. Zhou, X. Meng, and X. Zhou, "Role of inflammation, immunity, and oxidative stress in hypertension: New insights and potential therapeutic targets," Frontiers in Immunology, vol. 13, p. 1098725. [Online]. Available: https://doi.org/10.3389/fimmu.2022.1098725. [Accessed: Jan. 30, 2025].

[8] Centers for Disease Control and Prevention (CDC), "Antibiotic Use in the United States: Data and Statistics," CDC. [Online]. Available: https://www.cdc.gov/antibiotic-use/data-research/facts-stats/index.html. [Accessed: Jan. 30, 2025].

[9] World Health Organization, "Mental disorders," WHO. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/mental-disorders. [Accessed: Jan. 31, 2025].

[10] Greenhouse Treatment, "Long-Term Effects of Seroquel," Greenhouse Treatment. [Online]. Available: https://greenhousetreatment.com/what-is-seroquel/long-term-effects/. [Accessed: Jan. 31, 2025].