Performance Assessment 3 – Revision 1

D212 – Data Mining II

Jessica Hosey

MSDA, College of Information Technology

Western Governors University

November 8th, 2024

Part I: Research Question

A. 1. Propose one question relevant to a real-world organizational situation that you will answer using market basket analysis.

We are determining whether any prescriptions are purchased and associated with using Citalopram for patients in the hospital.

A. 2. Define one goal of the data analysis. Ensure your goal is reasonable within the scope of the selected scenario and is represented in the available data.

After completing counts of prescriptions ordered, there were a lot of heart medications ordered, and it is a hospital, so I did not question that. Most emergencies that a hospital would deal with are heart attacks and other heart-related emergencies that happen every day. However, I was more interested in medications used for anxiety as I suffer from Generalized Anxiety Disorder. Citalopram is a medication that is used to treat mood disorders like depression and some anxiety disorders. After some research, I found that it has been linked to causing heart problems in patients (Drugs.com 2024). This directed me to be concerned about (my research question) whether the heart medications ordered were at all correlated with patients who were ordered Citalopram as well or along with a heart-related medication.

Part II: Market Basket Justification

B. 1. Explain how the market basket analyzes the selected data set. Include expected outcomes.

Market Basket Analysis uses lists and analyzes them and creates "association rules" that apply to the dataset. For example, when purchasing items at a craft store, an individual may purchase a skein of yarn. The individual may also purchase other yarn items and a crochet hook to help them crochet something from the purchased yarn. Market Basket Analysis uses each transaction in the list to identify connections to other purchases in the dataset. Several statistics are computed about how many times yarn is purchased, as well as what, in addition to yarn, was also purchased at the store. This creates a rule or hypothesis: "If yarn is purchased, then crochet hooks are also purchased. Market Basket Analysis creates antecedents (the yarn) and consequents (crochet hooks, scissors, etc.). These antecedents and consequents are flipped, as are the rules. For my research question, we may see rules like "If Citalopram, then purchased X" or "If X, then purchased Citalopram."

B. 2. Provide one example of transactions in the data set.

Here is an example of a transaction in the dataset (see screenshot below). Each transaction has up to 20 spots available. Some transactions have only one or two prescriptions in the order so prescription 3-20 would contain a NaN or False value for it. This example is before I created a list of lists for the dataset and still identifies a transaction within the dataset.

```
#Show an example of a transaction within the dataset
df.iloc[0]
```

```
Presc01                    amlodipine
Presc02            albuterol aerosol
Presc03                  allopurinol
Presc04                 pantoprazole
Presc05                    lorazepam
Presc06                   omeprazole
Presc07                   mometasone
Presc08                  fluconozole
Presc09                   gabapentin
```

```
Presc10                  pravastatin
Presc11                       cialis
Presc12                     losartan
Presc13     metoprolol succinate XL
Presc14              sulfamethoxazole
Presc15                      abilify
Presc16               spironolactone
Presc17                albuterol HFA
Presc18                 levofloxacin
Presc19                 promethazine
Presc20                    glipizide
```

B. 3. Summarize one assumption of market basket analysis.

Market Basket Analysis's major assumption for the Apriori algorithm is that each purchase of items together (basket) has meaning. Those items typically purchased together are correlated, and the person buying them generally purchases them together. Therefore, putting these items near one another would be beneficial to the company in convincing shoppers to buy them together. For example, Amazon or Walmart (online grocery ordering) uses this to recommend or offer additional items that the shopper may have forgotten or would typically purchase with this item to increase ticket (order total) price.

Part III: Data Preparation and Analysis

C. 1. Transform the data set to make it suitable for market basket analysis. Include a copy of the cleaned data set.

See the attached files in my submission.

C. 2. Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate that the code is error-free.

```
[12]: #Using Apriori algorithm to generate frequent itemsets
      frequent_itemsets = apriori(clean_df, min_support = 0.02, use_colnames = True)
      frequent_itemsets
```

[12]:

	support	itemsets
0	0.046794	(Premarin)
1	0.238368	(abilify)
2	0.020397	(albuterol aerosol)
3	0.033329	(allopurinol)
4	0.079323	(alprazolam)
...	...	...
98	0.023064	(diazepam, lisinopril)
99	0.023464	(losartan, diazepam)
100	0.022930	(diazepam, metoprolol)
101	0.020131	(doxycycline hyclate, glyburide)
102	0.028530	(losartan, glyburide)

103 rows × 2 columns

```
[13]: #Use association_rules with lift of greater than 1
```

C. 3. Provide values for the association rules table's support, lift, and confidence.

```
#Use association_rules with lift of greater than 1
rules = association_rules(frequent_itemsets, metric = 'lift', min_threshold = 1.0)
rules
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(abilify)	(amlodipine)	0.238368	0.071457	0.023597	0.098993	1.385352	0.006564	1.030562	0.365218
1	(amlodipine)	(abilify)	0.071457	0.238368	0.023597	0.330224	1.385352	0.006564	1.137144	0.299568
2	(amphetamine salt combo)	(abilify)	0.068391	0.238368	0.024397	0.356725	1.496530	0.008095	1.183991	0.356144
3	(abilify)	(amphetamine salt combo)	0.238368	0.068391	0.024397	0.102349	1.496530	0.008095	1.037830	0.435627
4	(amphetamine salt combo xr)	(abilify)	0.179709	0.238368	0.050927	0.283383	1.188845	0.008090	1.062815	0.193648
...	...	...	...	...	...	...	...	...	...	...
89	(metoprolol)	(diazepam)	0.095321	0.163845	0.022930	0.240559	1.468215	0.007312	1.101015	0.352502
90	(doxycycline hyclate)	(glyburide)	0.095054	0.170911	0.020131	0.211781	1.239135	0.003885	1.051852	0.213256
91	(glyburide)	(doxycycline hyclate)	0.170911	0.095054	0.020131	0.117785	1.239135	0.003885	1.025766	0.232768
92	(losartan)	(glyburide)	0.132116	0.170911	0.028530	0.215943	1.263488	0.005950	1.057436	0.240286
93	(glyburide)	(losartan)	0.170911	0.132116	0.028530	0.166927	1.263488	0.005950	1.041786	0.251529

94 rows × 10 columns

C. 4. Explain the top three relevant rules generated by the Apriori algorithm. Include a screenshot of the top three relevant rules.

```
top_3_rules = rules[(rules['lift'] > 1.9) & (rules['confidence'] > 0.3)].sort_values(by=['lift'], ascending= False)
top_3_rules
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
75	(lisinopril)	(carvedilol)	0.098254	0.174110	0.039195	0.398915	2.291162	0.022088	1.373997	0.624943
73	(glipizide)	(carvedilol)	0.065858	0.174110	0.022930	0.348178	1.999758	0.011464	1.267048	0.535186
30	(metformin)	(abilify)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221

All three have a lift of 1.9 or higher and a confidence of 0.3 or higher. A higher lift indicates a strong correlation between the antecedent and the consequent, both included in the prescription orders. The confidence scores indicate that many transactions included either the antecedent or the consequent.

Part IV: Data Summary and Implications

D. 1. Summarize the analysis results' significance of support, lift, and confidence.

For support, the higher the number, the better the chances of patients purchasing the consequent (Gonzalez 2020). For our analysis (see screenshot below), all eight rules have a support value of around 0.02. This is not very high; however, some rules on the list are higher than others. Therefore, some rules regarding patients needing both medications during their visit may be accurate. There is a 2.1% support for patients who order Citalopram that will also need carvedilol, which is a heart-related medication

For lift, if the number is above 1, there is a high chance the patient will purchase the consequent. The opposite occurs if the number is below 1, and the patient does not buy the consequent. If the number is 1, then the antecedent does not affect the chance of purchasing the consequent (Gonzalez 2020). For our analysis (see screenshot below), all the rules have a lift above 1; therefore, there is an increased chance that patients will need the medication ordered. More specifically, patients who order Citalopram will also need to order amphetamine salt combo xr, which is used for treating narcolepsy and ADHD. There is a 1.41 lift value on the Citalopram carvedilol combo, which is concerning because Citalopram has been warned of heart-related issues during long-term usage.

For confidence, the higher the number, the better, and the higher the probability of the patient purchasing the consequent (Gonzalez 2020). For our analysis (see screenshot below), the confidence values range from 0.1 to 0.33. The rule with the highest confidence is the citalopram and amphetamine salt combo xr. There is a 24% chance that a patient who orders Citalopram must also order carvedilol.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
13	(citalopram)	(abilify)	0.087188	0.238368	0.024397	0.279817	1.173883	0.003614	1.057552	0.162275
45	(citalopram)	(amphetamine salt combo xr)	0.087188	0.179709	0.028796	0.330275	1.837830	0.013128	1.224818	0.499424
67	(citalopram)	(carvedilol)	0.087188	0.174110	0.021464	0.246177	1.413918	0.006283	1.095602	0.320707
81	(citalopram)	(glyburide)	0.087188	0.170911	0.021997	0.252294	1.476173	0.007096	1.108844	0.353384
12	(abilify)	(citalopram)	0.238368	0.087188	0.024397	0.102349	1.173883	0.003614	1.016889	0.194486
44	(amphetamine salt combo xr)	(citalopram)	0.179709	0.087188	0.028796	0.160237	1.837830	0.013128	1.086988	0.555754
66	(carvedilol)	(citalopram)	0.174110	0.087188	0.021464	0.123277	1.413918	0.006283	1.041163	0.354460
80	(glyburide)	(citalopram)	0.170911	0.087188	0.021997	0.128705	1.476173	0.007096	1.047650	0.389069

D. 2. Discuss the practical significance of your findings from the analysis.

Most doctors at the WGU facility are aware of the heart-related issues regarding long-term treatment with Citalopram, and they have their patient's care in mind while prescribing. Also, most patients who are on Citalopram are struggling with ADHD, Narcolepsy, Diabetes, and other issues. These issues could be a direct result of taking Citalopram, or it could be another issue. There is only one day of operation in this dataset; if there were more data points, other insights into this problem could be effectively analyzed.

D. 3. Recommend a course of action for the real-world organizational situation from part A1 based on the results from part D1.

Thankfully, only one heart-related medication was on the rules list for Citalopram for this single day of hospital operation. This could be something that the hospital may want to investigate with those patients to ensure long-term usage is not a factor.

The recommended course of action would be to continue gathering and analyzing data after many months of collection. The larger the dataset, the more reliable the analysis is and the more insights the company can provide its doctors.

Part V: Attachments

E. Panopto Video

See the attached files in my submission.

F. Sources

Drugs.com. (2024, February 29). Citalopram. Retrieved October 26, 2024, from https://www.drugs.com/citalopram.html

Gonzalez, D. (2020, October 27). Market basket analysis 101: Key concepts. Towards Data Science. https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00

KDnuggets. (2019, December 19). Market basket analysis: How to use it for data-driven decisions. KDnuggets. https://www.kdnuggets.com/2019/12/market-basket-analysis.html