

Raquel Ocasio
D212 – Data Mining II, Task 3
November 20, 2023
Western Governors University

Part A1

When a patient is prescribed a medication, which other medication(s) are typically prescribed with it?

Part A2

The goal of the data analysis is to use Market Basket Analysis to identify associations between medications that are prescribed to patients.

Part B1

The Market Basket Analysis technique analyzes the dataset by finding associations between prescribed medications. (Yolanda, 2023)

The Apriori algorithm was used in this Market Basket Analysis. The Apriori algorithm works by first finding associations within the dataset, and then pruning those associations which are infrequent. It then counts the frequency of each association, and discards those that do not meet a predefined minimum threshold. Finally, it uses these remaining associations to generate association rules, and a confidence value for each association rule.

The metrics used in Market Basket Analysis are support, confidence, lift, conviction, and leverage.

Support measures the frequency of occurrence of an itemset in the transaction database. It is calculated as the number of transactions containing the itemset divided by the total number of transactions.

Confidence measures the likelihood that an association rule is true. It is calculated as the support for the combined items in the rule divided by the support for the antecedent (left-hand side) itemset.

Lift measures the ratio of the observed support to the expected support if the items were independent. A lift value greater than 1 indicates that the items are positively correlated.

Conviction is a measure of how much more likely the consequent (right-hand side) of a rule is true, given that the antecedent is true. It is sensitive to the number of false positives.

Leverage measures the difference between the observed frequency of itemset occurrence and the frequency expected under independence. It ranges from -1 to 1, where values close to 1 indicate a strong association.

Expected outcome is a set of association rules. The association rules can be interpreted as an if-then rule, where if something is present, then something else is likely present or can happen.

Part B2

An example of a transaction in the dataset is below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
	Presc01	Presc02	Presc03	Presc04	Presc05	Presc06	Presc07	Presc08	Presc09	Presc10	Presc11	Presc12	Presc13	Presc14	Presc15	Presc16	Presc17	Presc18	Presc19	Presc20
	amlodipin	albuterol	allopurinol	pantopraz	lorazepam	omeprazole	mometasone	fluconazole	gabapentin	pravastatin	cialis	losartan	metoprolol	sulfamethoxazole	abilify	spironolactone	albuterol	levofloxacin	promethazine	glipizide

Part B3

One assumption of Market Basket Analysis is that each transaction is independent of others.

Part C1

See attached file.

Part C2

```
> # Run Apriori function on transactions dataset to uncover association rules
> arules <- apriori(basket, control = list(verbose = F), parameter = list(supp = 0.008, conf = 0.4, minlen = 2))
>
> # Remove redundant rules
> redundant_r <- is.redundant(arules)
> refined_arules <- arules[!redundant_r]
>
> # Inspect top rules sorted by parameter 'lift' in decreasing order
> inspect(head(sort(refined_arules, by = "lift", decreasing = T), 10))
```

	lhs	rhs	support
[1]	{lisinopril, metoprolol}	=> {carvedilol}	0.008665511
[2]	{amphetamine salt combo xr, lisinopril}	=> {carvedilol}	0.008932142
[3]	{atorvastatin, lisinopril}	=> {carvedilol}	0.009732036
[4]	{abilify, lisinopril}	=> {carvedilol}	0.017064391
[5]	{atorvastatin, metformin}	=> {abilify}	0.008532196
[6]	{citalopram, glyburide}	=> {amphetamine salt combo xr}	0.009065458
[7]	{lisinopril, metoprolol}	=> {abilify}	0.009198773
[8]	{amphetamine salt combo xr, lisinopril}	=> {abilify}	0.010131982
[9]	{diazepam, glipizide}	=> {abilify}	0.008265565
[10]	{atorvastatin, lisinopril}	=> {abilify}	0.011065191

```
confidence coverage lift count
[1] 0.5118110 0.01693108 2.939582 65
[2] 0.4466667 0.01999733 2.565426 67
[3] 0.4424242 0.02199707 2.541060 73
[4] 0.4169381 0.04092788 2.394681 128
[5] 0.5614035 0.01519797 2.355194 64
[6] 0.4121212 0.02199707 2.293265 68
[7] 0.5433071 0.01693108 2.279277 69
[8] 0.5066667 0.01999733 2.125563 76
[9] 0.5040650 0.01639781 2.114649 62
[10] 0.5030303 0.02199707 2.110308 83
>
> summary(refined_newrules)
set of 31 rules
```

Part C3

See attached file.

Part C4

For rule #1, if someone is prescribed lisinopril and metoprolol, they are 51% likely to be prescribed carvedilol as well. For rule #2, if someone is prescribed amphetamine salt combo xr and lisinopril, they are 45% likely to be prescribed carvedilol as well. For rule #3, if someone is prescribed atorvastatin and lisinopril, they are 44% likely to be prescribed carvedilol as well. See screenshot of the top three relevant rules below.

```
> # Inspect top rules sorted by parameter 'lift' in decreasing order
> inspect(head(sort(refined_arules, by = "lift", decreasing = T), 10))
```

	lhs	rhs	support
[1]	{lisinopril, metoprolol}	=> {carvedilol}	0.008665511
[2]	{amphetamine salt combo xr, lisinopril}	=> {carvedilol}	0.008932142
[3]	{atorvastatin, lisinopril}	=> {carvedilol}	0.009732036

	confidence	coverage	lift	count
[1]	0.5118110	0.01693108	2.939582	65
[2]	0.4466667	0.01999733	2.565426	67
[3]	0.4424242	0.02199707	2.541060	73

Part D1

Support indicates how frequently a set of items appear together in transactions. A high value for support means that the items are often appearing together. The results of the analysis showed that the support values were very low (less than .05) for each association, indicating that items did not frequently appear together.

Lift indicates how much more likely items are to be appear together than if they appeared separately. A lift value greater than 1 suggests a positive correlation, meaning the items are more likely to appear together. The results of the analysis showed that all the associations have a positive correlation, indicating that they are more likely to be prescribed together than separately.

Confidence measures how likely item B is prescribed when item A is prescribed. The results of the analysis show for all the association rules, the confidence in the association was 56% or less. Only seven of the association rules had a confidence value of 50% or more. This indicates that the majority of the association rules were likely to happen.

Part D2

These results for the Support or Confidence values do not appear to have practical significance as they do not indicate a strong association between the prescribed medications. The results for the Lift values appear to have practical significance, since it provides insight as to which medications are more likely to be prescribed together than separately.

Part D3

The stakeholders can use the results of this analysis to further research medication prescriptions and how they affect patient admission or re-admission.

Part E

The demonstration can be viewed at

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=c95c71fa-c62b-41b5-8772-b0c001732775>

Part F

Yolanda, R. (2023, August 21). Market Basket Analysis with R - Rizka Yolanda - Medium.

Medium. <https://medium.com/@yolandawiyono98/market-basket-analysis-with-r-8001417a8e29>