OFM4 Task 3: Association Rules and Lift Analysis

Steven Oldenburg

Western Governors University

D212: Data Mining II

04/1/2024

OFM4 Task 3: Association Rules and Lift Analysis

A1: Proposal of Question

What is the most frequently purchased item in the dataset and can we use lift analysis to measure a significant association beyond random chance with another item?

A2: Defined Goal

The goal of the analysis is to provide potential associations between the most frequently purchased item and other items in the dataset. Lastly, this will be followed up with a lift analysis to show which items are purchased together beyond random chance. The executives asked for information regarding items customers would most likely be interested in an effort to reduce churn rates. The goal is to identify the product most customers likely purchased and the associated purchases to create information for our marketing department.

B1: Explanation of Market Basket

A market basket analysis works by using algorithms like the Apriori algorithm to identify item combinations that happen often together. It does this by measuring the strengths of associations between the items, such as how often a customer buys a flashlight and buying batteries. The expected outcome and goal of this analysis is to uncover hidden patterns by quantifying the associations between the items. These patterns may be less obvious than batteries and flight lights, such as which fiction novels people who enjoy sci-fi might also purchase. It's most frequently used in upselling or cross-selling. For instance, Disneyland might offer a credit card, but to find the best way to upsell their credit card at the cash register they might do a market analysis on which characters are purchased the most, then offer a credit card with those characters instead of randomly selecting which characters to place on the credit card.

B2: Transaction Example

Transaction Example #39: iPhone 12 case, USB Type-C Cable, Apple USB-C Charger Cable

B3: Market Basket Assumption

The main assumption of a market-based analysis is that customers who purchase items together are likely to purchase the items together again in the future or are interested in items that are often purchased together with those items. The assumption does guarantee a relationship between the two and this assumption is stronger in items that are more frequently purchased together. The main idea is that that preference or a situation itself could be the reason for buying items and a pattern may emerge. For instance, one customer may like cakes, so they buy cake mix and sugar while the next may have a birthday cake, so they buy cake mix, sugar, and candles. The logical assumption in this scenario is that people who buy cake might often buy sugar or vice versa, regardless of the reason.

C1: Transforming the Data Set

I've attached the transformed dataset as "clean_df.csv".

C2: Code Execution

```
#Frequent Items
frequent_itemsets = apriori(clean_df, min_support = 0.05, use_colnames = True)
#Load main rules
rules = association_rules(frequent_itemsets, metric = 'lift', min_threshold = 1.0)
rules
```

[22]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(dust-off compressed gas 2 pack)	(10ft iphone charger cable 2 pack)	0.238368	0.050527	0.023064	0.096756	1.914955	0.011020	1.051182	0.627330
1	(10ft iphone charger cable 2 pack)	(dust-off compressed gas 2 pack)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
2	(dust-off compressed gas 2 pack)	(anker usb c to hdmi adapter)	0.238368	0.068391	0.024397	0.102349	1.496530	0.008095	1.037830	0.435627
3	(anker usb c to hdmi adapter)	(dust-off compressed gas 2 pack)	0.068391	0.238368	0.024397	0.356725	1.496530	0.008095	1.183991	0.356144
4	(anker usb c to hdmi adapter)	(vivo dual lcd monitor desk mount)	0.068391	0.174110	0.020931	0.306043	1.757755	0.009023	1.190117	0.462740
...	...	...	...	...	...	...	...	...	...	...
89	(vivo dual lcd monitor desk mount)	(screen mom screen cleaner kit)	0.174110	0.129583	0.035462	0.203675	1.571779	0.012900	1.093043	0.440468
90	(stylus pen for ipad)	(usb 2.0 printer cable)	0.095054	0.170911	0.020131	0.211781	1.239135	0.003885	1.051852	0.213256
91	(usb 2.0 printer cable)	(stylus pen for ipad)	0.170911	0.095054	0.020131	0.117785	1.239135	0.003885	1.025766	0.232768
92	(stylus pen for ipad)	(vivo dual lcd monitor desk mount)	0.095054	0.174110	0.025197	0.265077	1.522468	0.008647	1.123778	0.379218
93	(vivo dual lcd monitor desk mount)	(stylus pen for ipad)	0.174110	0.095054	0.025197	0.144717	1.522468	0.008647	1.058066	0.415518

94 rows × 10 columns

C3: Association Rules Table

Below is the table for the rubric:

```
selected_columns = ['antecedents', 'consequents', 'support', 'lift', 'confidence']
lift_groups = rules[selected_columns]
lift_groups
```

	antecedents	consequents	support	lift	confidence
0	(dust-off compressed gas 2 pack)	(10ft iphone charger cable 2 pack)	0.023064	1.914955	0.096756
1	(10ft iphone charger cable 2 pack)	(dust-off compressed gas 2 pack)	0.023064	1.914955	0.456464
2	(dust-off compressed gas 2 pack)	(anker usb c to hdmi adapter)	0.024397	1.496530	0.102349
3	(anker usb c to hdmi adapter)	(dust-off compressed gas 2 pack)	0.024397	1.496530	0.356725
4	(anker usb c to hdmi adapter)	(vivo dual lcd monitor desk mount)	0.020931	1.757755	0.306043
...	...	...	...	...	...
89	(vivo dual lcd monitor desk mount)	(screen mom screen cleaner kit)	0.035462	1.571779	0.203675
90	(stylus pen for ipad)	(usb 2.0 printer cable)	0.020131	1.239135	0.211781
91	(usb 2.0 printer cable)	(stylus pen for ipad)	0.020131	1.239135	0.117785
92	(stylus pen for ipad)	(vivo dual lcd monitor desk mount)	0.025197	1.522468	0.265077
93	(vivo dual lcd monitor desk mount)	(stylus pen for ipad)	0.025197	1.522468	0.144717

94 rows × 5 columns

C4: Top 3 Rules

```
top_3 = rules[(rules['lift'] > 1.2) & (rules['support'] > .05) & (rules['confidence'] > 0.25)].sort_values(by=['lift'], ascending= False)
top_3
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
4	(dust-off compressed gas 2 pack)	(vivo dual lcd monitor desk mount)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008	0.400606
5	(vivo dual lcd monitor desk mount)	(dust-off compressed gas 2 pack)	0.174110	0.238368	0.059725	0.343032	1.439085	0.018223	1.159314	0.369437
2	(hp 61 ink)	(dust-off compressed gas 2 pack)	0.163845	0.238368	0.052660	0.321400	1.348332	0.013604	1.122357	0.308965

All three rules have a support of greater than .05 which means each represent a number greater than 5% of the total transactions. All three also have a lift value above 1.3 showing a significant probability beyond random chance that these items will be purchased together.

Rule 1: vivo dual lcd monitor desk mount -> dust-off compressed gas 2 pack

Confidence (0.343032): This is the highest confidence value among the three rules. If a transaction contains a vivo dual lcd monitor desk mount, there's a 34.3% chance of it also containing dust-off compressed gas 2 pack.

Rule 2: dust-off compressed gas 2 pack -> vivo dual lcd monitor desk mount

Confidence (0.250559): This has a confidence measurement of 25%. While the implication is weaker, it still means that if someone buys a dust off compressed gas 2 pack first, 1 out of every 4 will purchase the mount as well.

Rule 3: hp 61 ink -> dust-off compressed gas 2 pack

Confidence (0.321400): This has the second-highest confidence, suggesting one out of every 3 transactions of finding "dust-off compressed gas 2 pack" with "hp 61 ink" purchases.

Based on all of these values, these three items are items our customers actively seek out.

D1: Significance in Analysis

In our frequent items analysis, we found that the most frequent item is dust-off compressed gas which comes up in almost 6% based on the support of .059 of all transactions and it is most often purchased with a vivo dual LCD monitor mount. A confidence score of .34 suggests that customers who buy the desk mount also buy dust off 34% of the time. This transaction has the highest lift of 1.439 which shows how often when one is purchased, so is another beyond random chance. Any lift amount above 1 shows a positive association, so this metric shows this rule happens often. The last rule we found is that ink for an HP 61 is often purchased first when the same dust-off gas is second. This is found in nearly the same support of .052 and a lift of 1.348 with confidence of .32 suggesting that about a third of the transactions with ink also have dust off being purchased. This information can be summarized and simplified as our customers purchase a lot of dust off, especially since it's found in multiple rules. It is by far our most popular product amongst our customers and provides the best opportunities for the company's campaigns to cross-sell or reduce churn rates through discounts.

D2: Practical Significance

The practical significance of these findings suggests that customers who buy a new monitor mount often buy dust off to clean their working stations around a third of the time. We also found customers who buy ink for printers tend to purchase dust off a third of the time. Together this suggests that customer who plan to spend money upgrading their computer stations or maintain a steady supply of ink are more interested in dusting off. Customers who also purchase dust off are interested in a new monitor mount, but only one-fourth of the time. We these findings we can use them to cross-sell to raise revenue or target specific markets to try and reduce churn rates such as offering discounts for items.

D3: Course of Action

I would suggest setting up a task force that includes marketing, regional managers, and product managers to discuss the best way to apply this data to lower churn rates. Based on the association rule analysis, a significant portion of customers who purchase dust-off compressed gas also tend to buy monitor mounts. Offering a free monitor mount for customers who renew their contracts and offering discounts on dust-offs for current customers would be a great incentive to stay with our company based on our customer's habits. This, however, needs to be reviewed by product management to determine the cost it would have to create this marketing strategy. The idea is to offer a large item customers are interested in to renew contracts which will lead to lower churn rates and offer a discount on dust off which our customers often use regularly to try and keep our customers loyal or to at least have a continued benefit of value our customers would often use. If management decides to go forward with this initiative, we will require feedback from all stakeholders on the best way to proceed and how to market this initiative to our customer base. I also suggest A/B testing on which combinations would provide the most value to our company and our target customers. Lastly, we can explore the possibility of loyalty rewards for items such as dust off as an alternative approach.

Third-Party Code

No third-party code was used.

References