Report

By: Tyler McElroy

Market basket analysis is used to discover associations between products that customers tend to purchase together. By analyzing customer transaction data, it reveals which items are frequently bought as a group, helping businesses understand purchasing patterns and implement various strategies to increase revenue.

To perform market basket analysis, I used the apriori algorithm via the mlxtend library in Python. The apriori algorithm identifies associations between products using several metrics: (1) confidence, (2) lift, and (3) support. Refer below for an in-depth description for each metric.

1.) Confidence:

Measures how likely B is to occur given A. Interpreted as the probability of B, assuming A has occurred.

$$\text{Confidence}(A \to B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)} = \frac{\text{Transactions with A and B}}{\text{Transactions with A}}$$

2.) Lift:

Measures how much more likely A and B occur together than if they were statistically independent. Lift > 1 means A and B are positively associated and are good candidates for cross-selling.

$$\text{Lift}(A \to B) = \frac{\text{Confidence}(A \to B)}{\text{Support}(B)} = \frac{\text{Support}(A \cup B)}{\text{Support}(A) \times \text{Support}(B)}$$

3.) Support:

Measures how frequently A and B appear together in the dataset. Typically used as a threshold to filter out rare rules.

$$\text{Support}(A \to B) = \frac{\text{Number of transactions containing both A and B}}{\text{Total number of transactions}}$$

The dataset used to perform market basket analysis is a synthetic dataset representing online grocery purchases. More specifically, the dataset contains grocery purchases of customers via an online portal, like the ones used by Target, Walmart, and Amazon. By seeing what customers have purchased previously, it allows for companies to cross-sell through recommendations via their online portal.

After performing market basket analysis on the dataset, the following combinations were listed as important using specific thresholds for confidence, lift, and support.

	antecedents	consequents	support	confidence	lift
67	(bread, cereal)	(jam)	0.025208	0.363077	1.230676
29	(apples, spinach)	(chicken)	0.024354	0.331395	1.223392
23	(apples, pasta)	(cheese)	0.024995	0.334286	1.220586
108	(butter, pasta)	(spinach)	0.025636	0.336134	1.217837
49	(butter, beans)	(oranges)	0.025636	0.338983	1.211282
116	(cheese, jam)	(cereal)	0.025422	0.319892	1.188426
62	(rice, beans)	(oranges)	0.024995	0.332386	1.187710
178	(jam, yogurt)	(oranges)	0.026704	0.328084	1.172337
165	(rice, eggs)	(oranges)	0.022431	0.327103	1.168831
57	(beans, jam)	(oranges)	0.026063	0.326203	1.165616
109	(butter, spinach)	(pasta)	0.025636	0.326975	1.161284
141	(pasta, yogurt)	(cheese)	0.024995	0.316216	1.154609
140	(cheese, yogurt)	(pasta)	0.024995	0.325000	1.154268
47	(oranges, butter)	(beans)	0.025636	0.323450	1.154017
148	(oranges, chicken)	(rice)	0.023072	0.318584	1.153358
134	(cheese, jam)	(eggs)	0.023927	0.301075	1.152358
187	(oranges, pasta)	(rice)	0.025208	0.318059	1.151458
186	(oranges, rice)	(pasta)	0.025208	0.324176	1.151341
117	(cereal, jam)	(cheese)	0.025422	0.314815	1.149492
92	(butter, cheese)	(rice)	0.022858	0.317507	1.149460
119	(pasta, cereal)	(cheese)	0.022858	0.314706	1.149094
123	(milk, cereal)	(chicken)	0.022645	0.310850	1.147548
60	(oranges, rice)	(beans)	0.024995	0.321429	1.146804
188	(rice, pasta)	(oranges)	0.025208	0.320652	1.145781
6	(apples, bread)	(cheese)	0.024781	0.313514	1.144740

The threshold used for lift is any value greater than 1.1; the threshold for confidence is any value equal to or greater than 0.3; and the threshold value for support is any value greater than or equal to 0.01. In practice, this means that combinations are selected which contain a stronger chance than random at being true, at least a 30% chance of a follow-up purchase via the consequent, and the combinations were seen in at least 1% of transactions.

As for testing the efficacy of the combinations listed above, rules can be implemented into the application customers use to order groceries online. A/B testing can be performed for the added application functionality to evaluate the rules. Further refinement can be performed once the results from A/B testing are obtained.