Analyzing Grocery Market Baskets and Recipes

Team:
Jadyn Ellis, Christine Widden, Alex Arrieta,
Ryuhei Shida

Goals of Analysis

Ultimate Goal: Build a recommender engine for additional item purchases based on current basket

- **Subproblem 1:** Association Rules
  - Discover what ingredients tend to be used together and what ingredients imply others

- **Subproblem 2:** Clustering
  - Find common groupings of items in grocery purchases

The Data: Sourced from Kaggle

- Food.com
  - Contains information pertaining to *ingredients for a given recipe*
  - 8000 unique food id's over 178,000 unique recipes

- Instacart
  - Contains information pertaining to *items purchased in grocery orders*
  - Over 3 million market baskets with almost 50,000 unique items

The Data: Cleaning

- Food.com
  - Originally contained extraneous information about the recipe
  - Cleaned to provide a list of ingredients on each row

1	389	7655	6270	1527	3406							
2	2683	4969	800	5298	840	2499	6632	7022	1511	3248	4964	6270

The Data: Cleaning

- Instacart
  - Had too many overly specific items like “Robust Golden Unsweetened Oolong Tea” and non-food items
  - Matched to the same ID system as the Food.com dataset

840	7625	715	5010	3670	5185	5010
837	5205	150	2145			

  - Each row represents a single Instacart grocery order
  - Ids represent a unique food item

Subproblem 1: Association Rules

Methodology:

- Apriori Algorithm
  - *If X is a frequent itemset in T, then every non-empty subset of X is also a frequent itemset in T*
- Skyline Frequent Itemsets
- Only one item on right side of association rule
  - Permutation Properties

Subproblem 1: Association Rules

Hyperparameter Values

- **Minimum support: 0.01**
  - Achieves a larger set of frequent itemsets for more information
- **Minimum Confidence: 0.7**
  - Achieves a set of moderate-strong relationships while also resulting in larger set of association rules for more information

Subproblem 1: Association Rules

Highest Confidence: 0.8858

{butter, baking powder, baking soda} → {egg}

Highest Support: 0.0358

{onion, pepper} → {salt}

Subproblem 1: Association Rules

Tendencies of Rules:

- Lots of rules were associated with ingredients commonly found in baked goods, such as {egg, baking soda, all-purpose flour} → {salt} or {vanilla, sugar, flour} → {egg}
  - Logically makes sense since a lot of baking recipes include the same base of ingredients
  - Savory dishes tend to be more unique

- May contain biases towards baking associations in our final recommendation system

Subproblem 2: Clustering

Methodology:

- Hierarchical Clustering
- Random subset
  - 3,000,000 rows = computational restrictions
- Cutoff value: 0.8
- Distance metric: Dice
  - Ranges from 0 to 1
  - $$D(P, Q) = \frac{2|P \cap Q|}{|P| + |Q|}$$

Subproblem 2: Clustering

Results: 40 total clusters of grocery orders

Found similarities in larger clusters

- Water and Fruit items
- Avocado, Ice, Fats
- Gluten, Broccoli, Coconuts, Almonds

Conclusions: Not helpful in developing informative similarities between grocery orders

Putting it all Together: Recommendation System

Methodology:

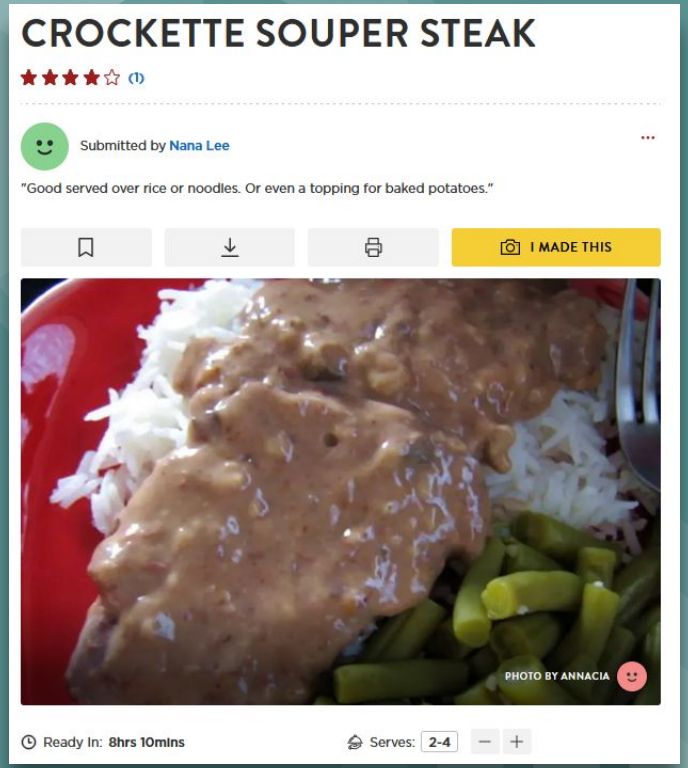
- Two methods of making recommendations, purchase recommendation and recipe recommendation
- Informed by information retrieval methodologies
- Purchase recommendation uses association rules to identify new items to purchase
- Recipe recommendation uses a simple % overlap similarity measure

Putting it all Together: Recommendation System

Given a list of items, we can recommend recipes that the user may want to try. With a list of ingredients, we look at every recipe and identify each one that we have 60% or more of the ingredients for, returning that recipe's name.

Given the list of ingredients salt, pepper, and steak, we can recommend these two recipes from Food.com:

- "Crockette Souper Steak", which calls for pepper, cream of mushroom soup, steak, and salt
- "Sweet Vermouth Country Style Ribs", which call for pepper, pork ribs, and salt



Putting it all Together: Recommendation System

Given the market basket of milk, water, butter, eggs, and sugar, the following association rules and recommended items are produced:

This will lead us to recommend this user additionally purchases garlic cloves, salt, soy sauce, cinnamon, and all-purpose flour

Association Rules	
Left	Right
{'water'}	{'soy sauce'}
{'sugar'}	{'garlic clove'}
{'milk'}	{'water'}
{'butter'}	{'garlic clove'}
{'sugar', 'butter'}	{'cinnamon'}
{'butter', 'milk'}	{'all-purpose flour}
{'water', 'butter'}	{'salt'}
{'water', 'sugar'}	{'salt'}
{'sugar', 'butter', 'milk'}	{'salt'}