

Building a Recipe Recommender

Flora Xinru Cheng

Natural Language Processing and Unsupervised Learning Project

Motivation

“Food is our common ground, a universal experience.”

—James Beard (1903-1985)

Motivation

“Food is our common ground, a universal experience.”

—James Beard (1903-1985)

Motivation

“Food is our common ground, a universal experience.”

—James Beard (1903-1985)

“Try them! You may like them! ”

Motivation

“Food is our common ground, a universal experience.”

—James Beard (1903-1985)

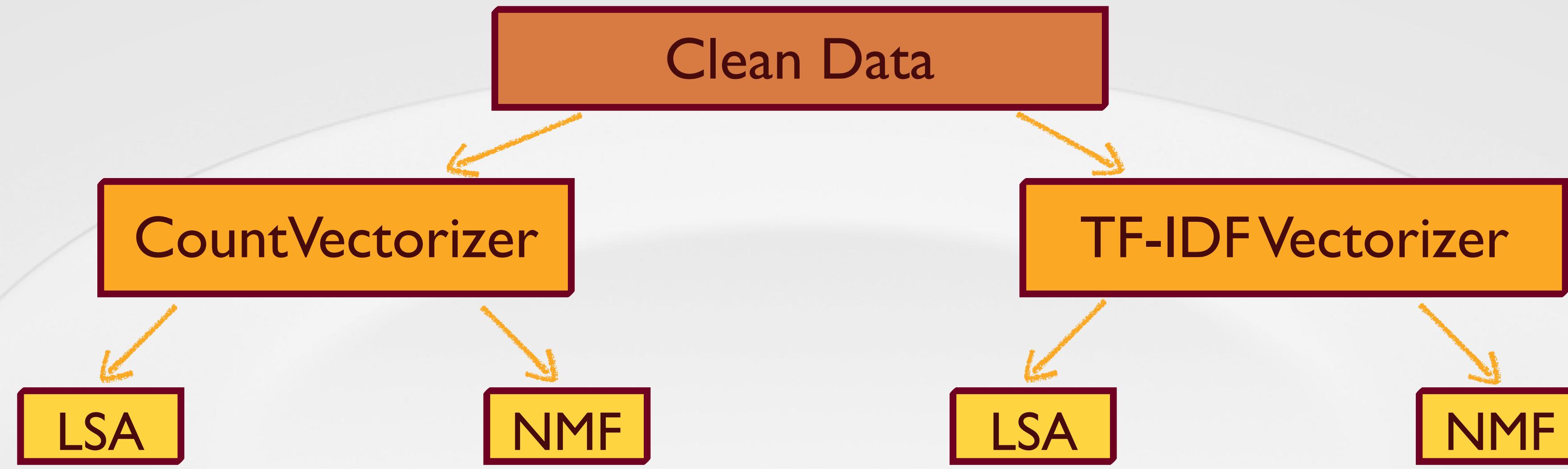
“Try them! You may like them! ”

Motivation

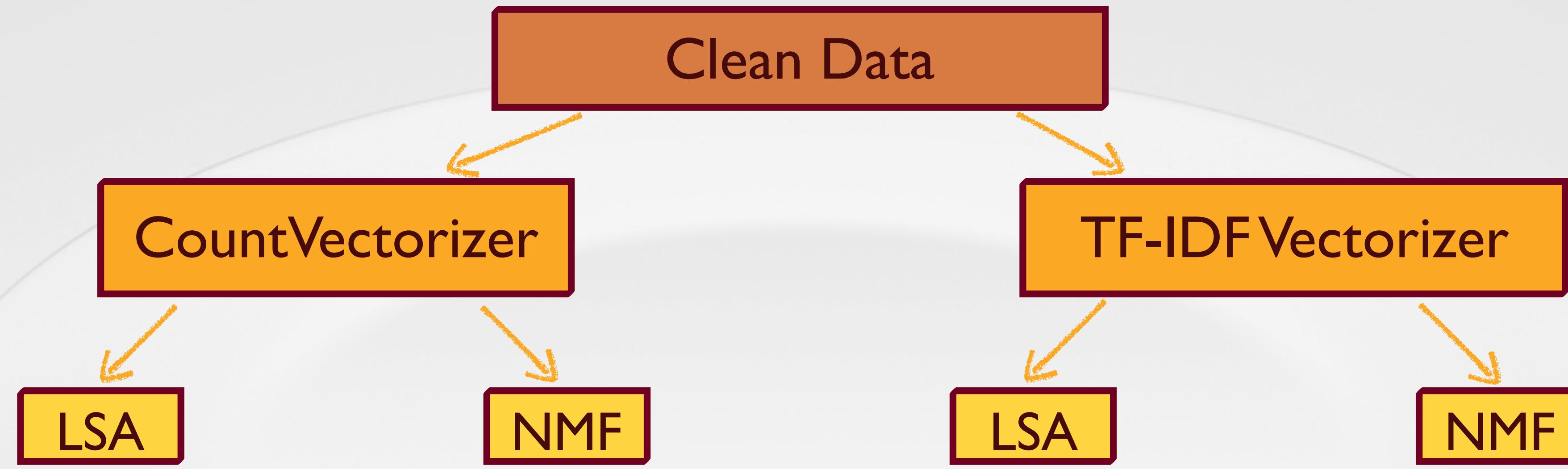


- * RecipeBox (~125,000 scraped recipes from Allrecipes.com, Epicurious.com, Foodnetwork.com)
 - * 50.9MB, in JSON format (title, ingredients, instructions)
- * Kaggle What's Cooking dataset (ingredients, cuisine types)

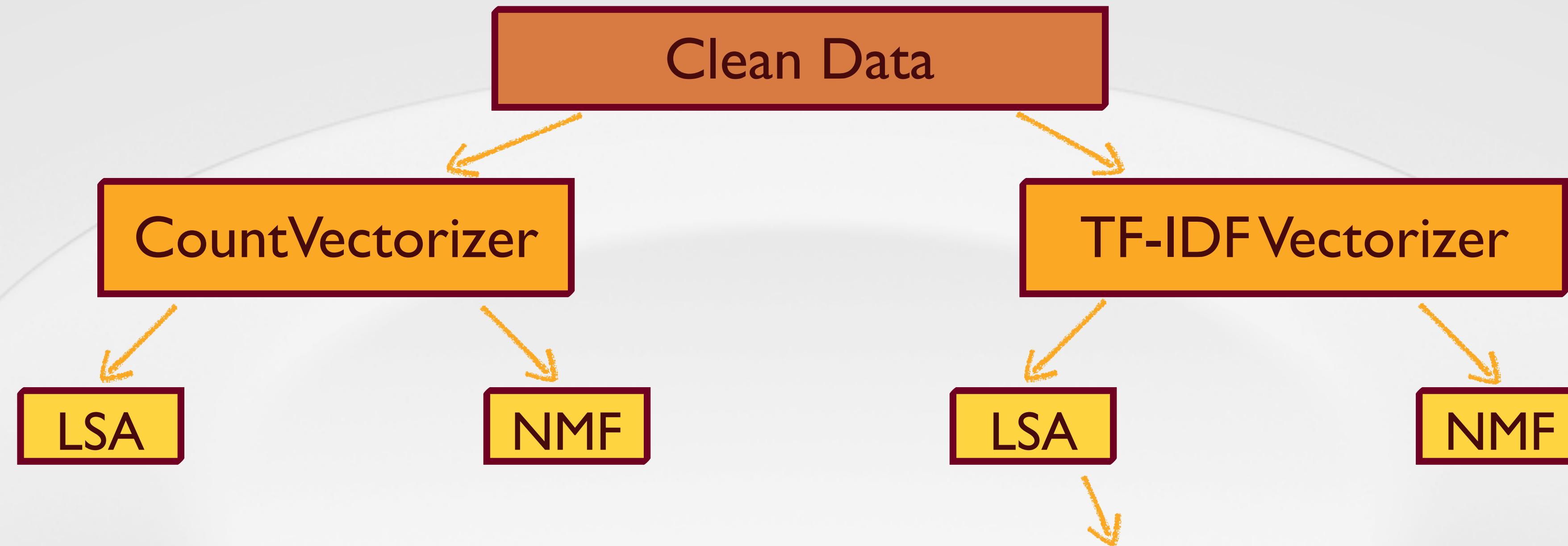
Data



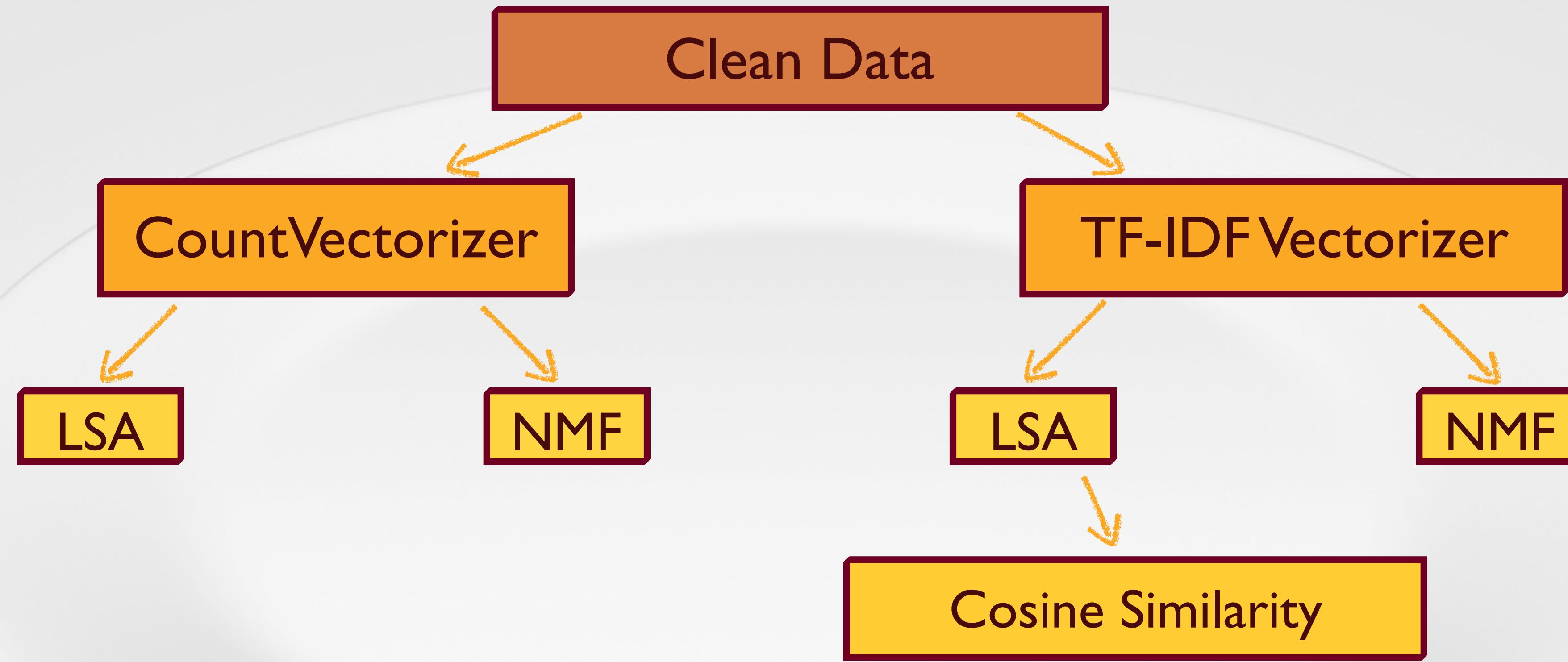
Methodology



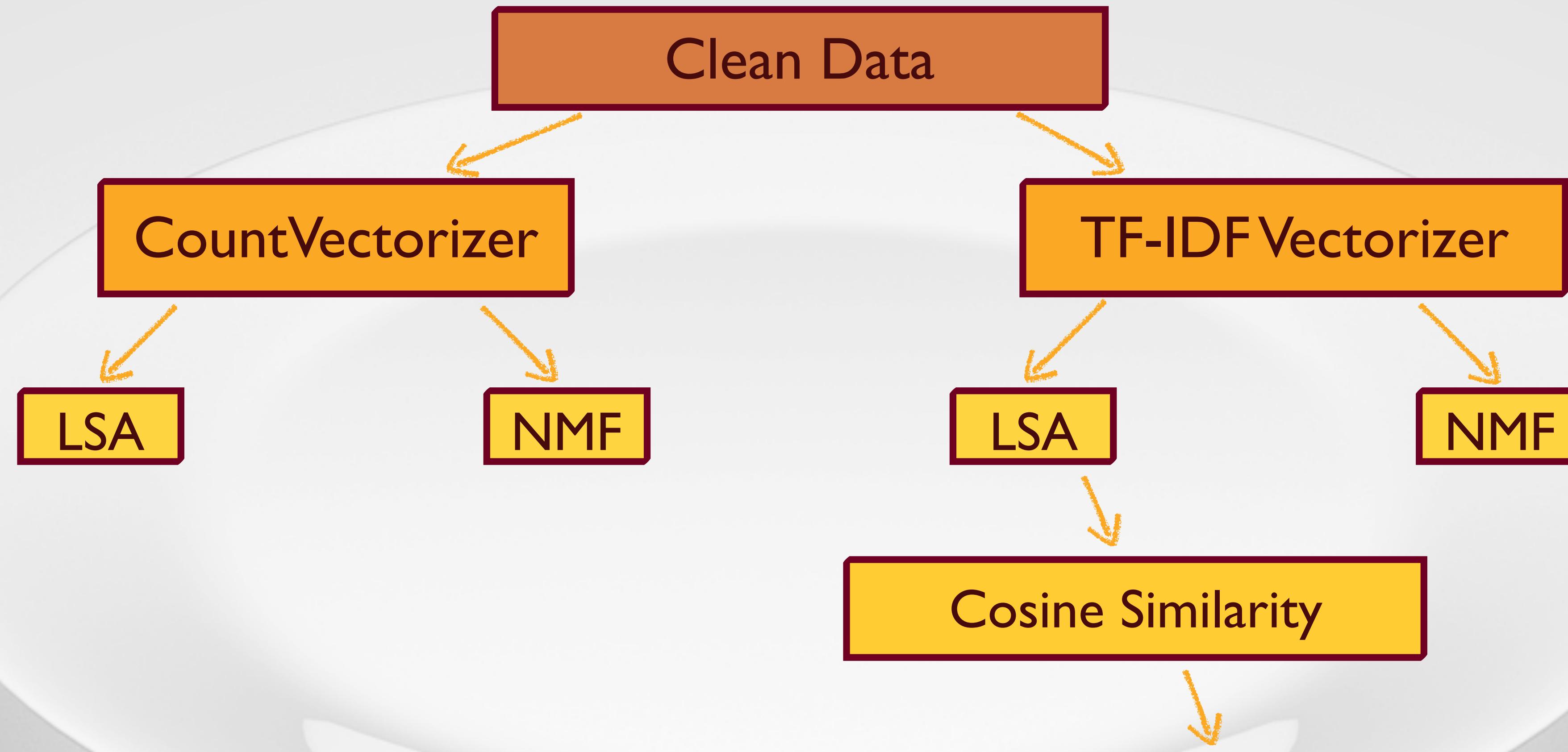
Methodology



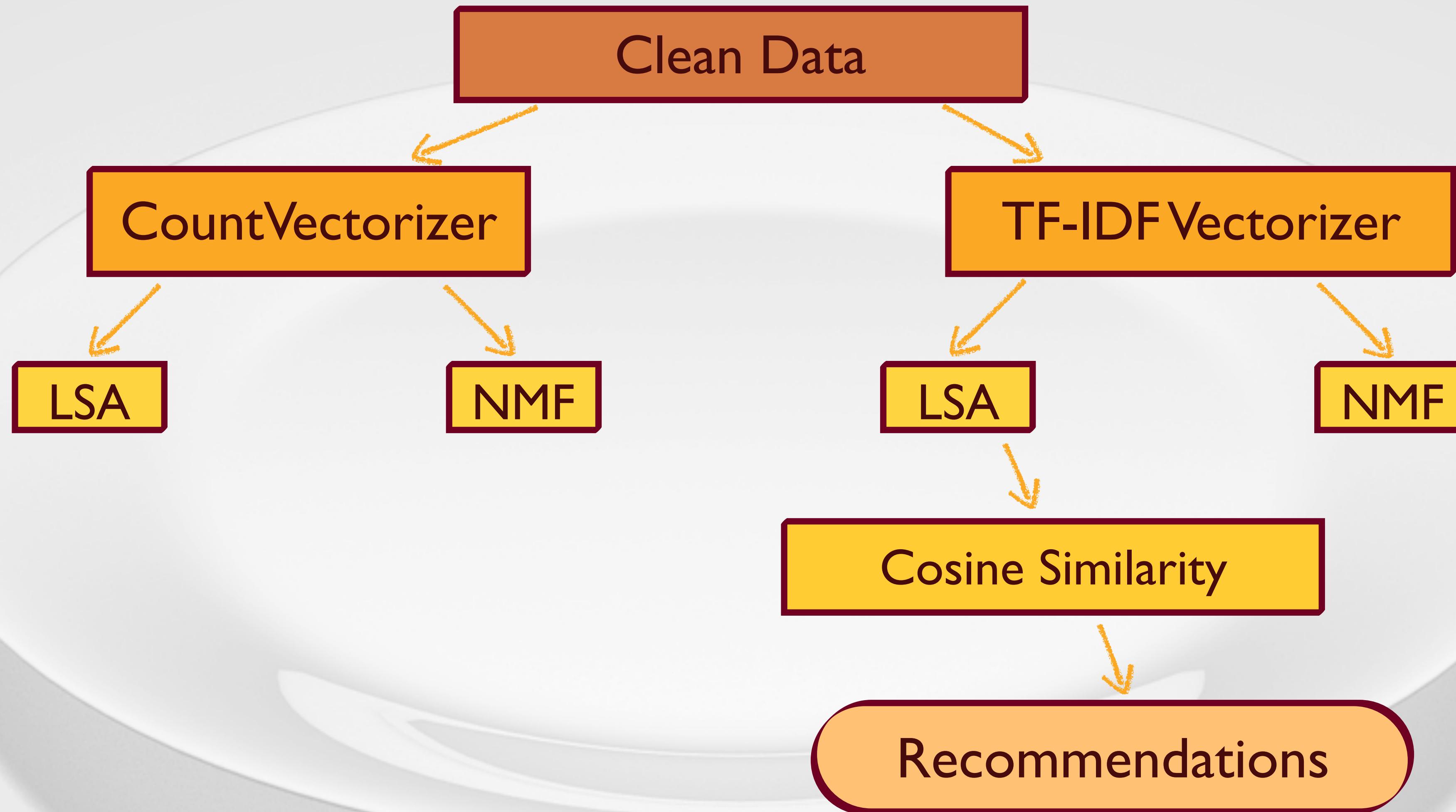
Methodology



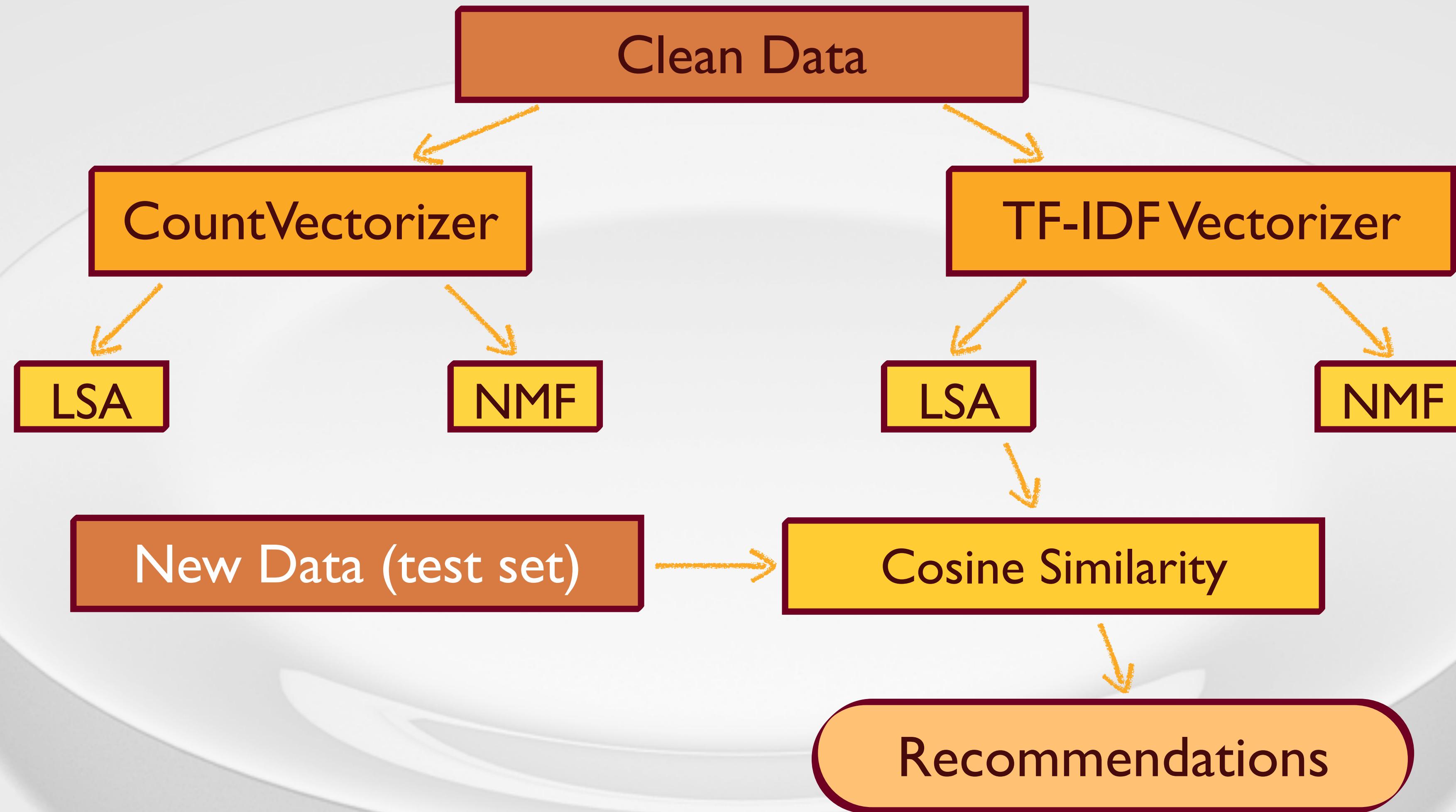
Methodology



Methodology



Methodology



Methodology

* Recommendations (TF-IDF, LSA):



Best Chocolate Chip Cookies



Edible Chocolate Chip Cookie Dough

Golden Chocolate Chip Cookies

Caramel Pecan Cookies

Giant Chocolate Chip Cookie

Results

* Recommendations (TF-IDF, LSA):

Awesome Slow Cooker Pot Roast



Results

Beef Tips and Noodles

Slow Cooker London Broil

Pot Roast in Foil

Polish Chop Suey

Home > Recipes > Main Dish > Beef > Pot Roast

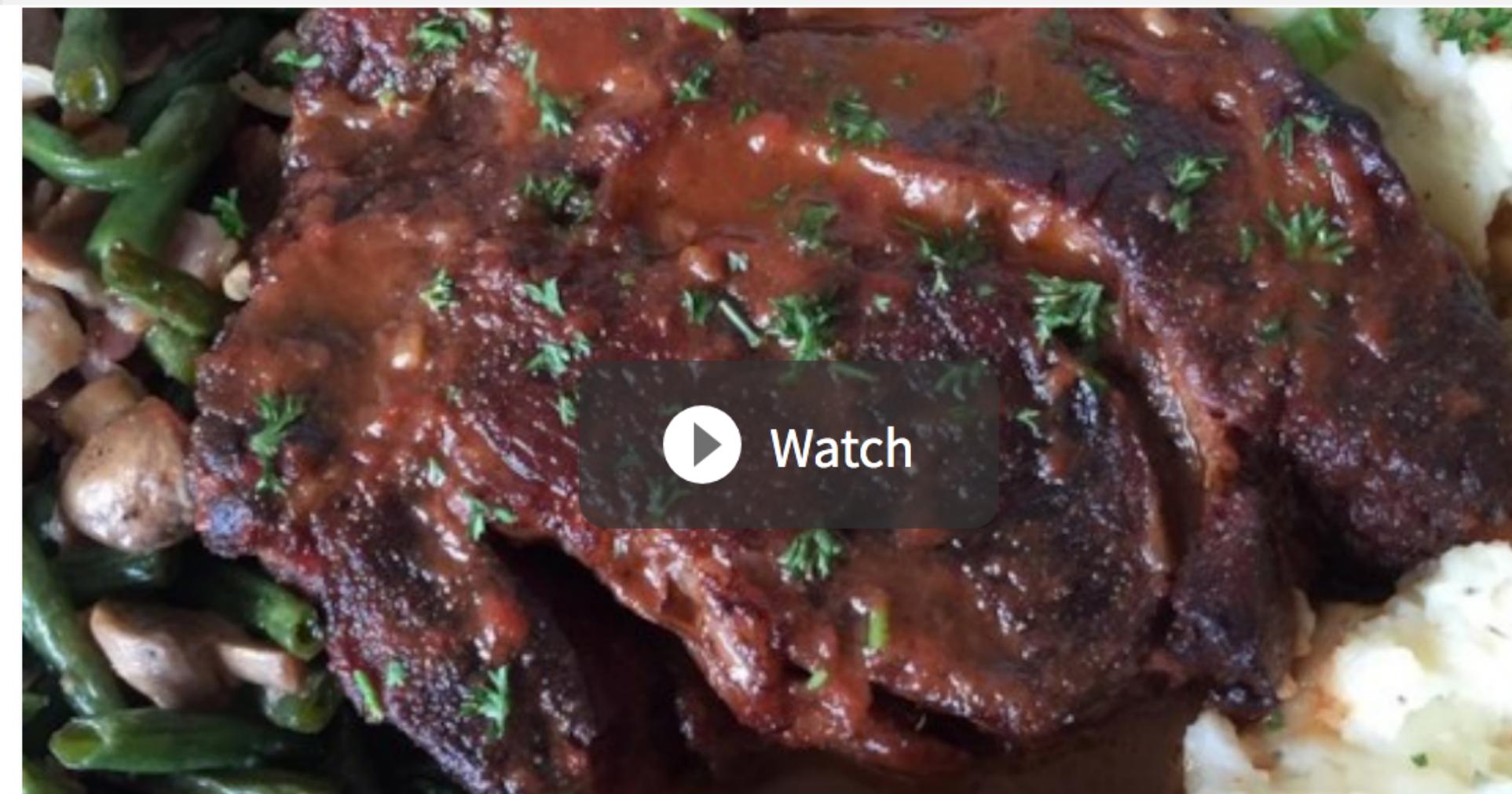
Awesome Slow Cooker Pot Roast



10k made it | 7k reviews | 207 photos

Recipe by: Brenda Arnold

"This is a very easy recipe for a delicious pot roast. It makes its own gravy. It's designed especially for the working person who does not have time to cook all day, but it tastes like you did. You'll want the cut to be between 5 and 6 pounds."



Awesome Slow Cooker Pot Roast



Results

Slow Cooker London Broil

Pot Roast in Foil

Polish Chop Suey

Recommended

See more



London Broil for the Slow Cooker

326

By Yvonne



Slow Cooker Steak Fajitas

13

By Miguel Ruiz

It works!

allrecipes

BROWSE ▾ Find a recipe Ingredient Search     Create a profile 

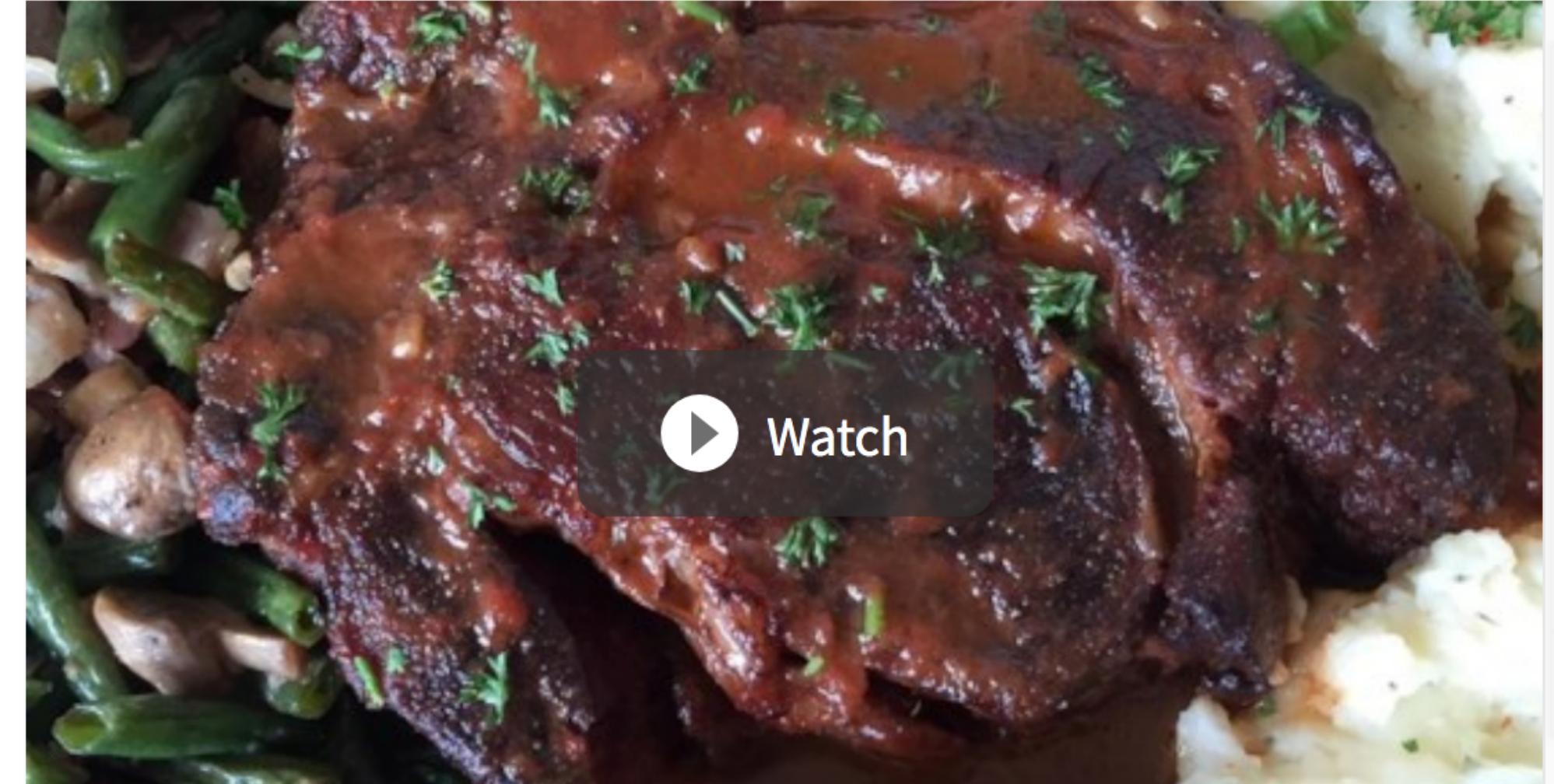
Home > Recipes > Main Dish > Beef > Pot Roast

Awesome Slow Cooker Pot Roast

★★★★★ 5 stars
10k made it | 7k reviews | 207 photos

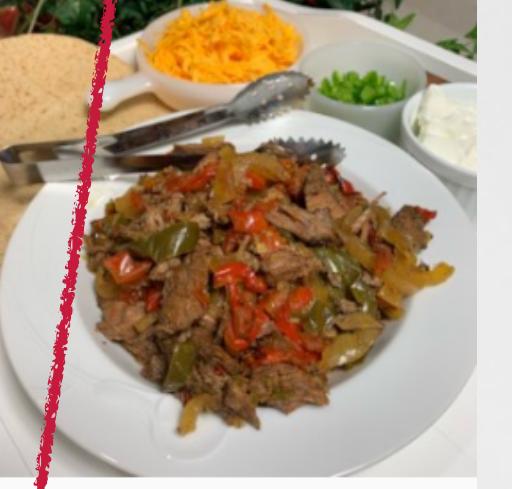
Recipe by: Brenda Arnold

"This is a very easy recipe for a delicious pot roast. It makes its own gravy. It's designed especially for the working person who does not have time to cook all day, but it tastes like you did. You'll want the cut to be between 5 and 6 pounds."



Recommended


London Broil for the Slow Cooker
★★★★★ 326 By Yvonne


Slow Cooker Steak Fajitas
★★★★★ 13 By Miguel Ruiz

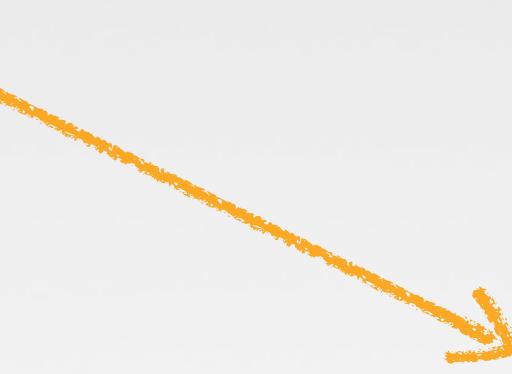
Awesome Slow Cooker Pot Roast

Results

Slow Cooker London Broil
Pot Roast in Foil
Polish Chop Suey

* Recommendations (TF-IDF, LSA):

Jalapeño and Lime-Marinated Skirt Steak Tacos



Green Bean and Bass Pouches

Nachos My Way

Tequila-Lime Mahimahi Tacos

Coconut Red Curry Sauce and Noodles

Results



3.2 Recommendation on new data

3.2.1 Small test data

```
In [116]: testdata=pd.DataFrame(columns = ["title","ingredients"])
executed in 12ms, finished 11:45:07 2019-11-14
```

```
In [117]: inglist1=[ ''
 2 tablespoons reduced sodium soy sauce
 2 tablespoons freshly squeezed lime juice
 2 tablespoons canola oil, divided
 3 cloves garlic, minced
 2 teaspoons chili powder
 1 teaspoon ground cumin
 1 teaspoon dried oregano
 1 1/2 pounds skirt steak, cut into 1/2-inch pieces
 12 mini flour tortillas
 3/4 cup diced red onion

 1/2 cup chopped fresh cilantro leaves
 1 lime, cut into wedges
 ''
]
title1='Mexican Street Tacos by DamnDelicious'
```

Demonstration

3.2 Recommendation on new data

3.2.1 Small test data

```
In [116]: testdata=pd.DataFrame(columns = ["title","ingredients"])
executed in 12ms, finished 11:45:07 2019-11-14
```

```
In [117]: inglist1=[ ''
 2 tablespoons reduced sodium soy sauce
 2 tablespoons freshly squeezed lime juice
 2 tablespoons canola oil, divided
 3 cloves garlic, minced
 2 teaspoons chili powder
 1 teaspoon ground cumin
 1 teaspoon dried oregano
 1 1/2 pounds skirt steak, cut into 1/2-inch pieces
 12 mini flour tortillas
 3/4 cup diced red onion

 1/2 cup chopped fresh cilantro leaves
 1 lime, cut into wedges
 ''
]
title1='Mexican Street Tacos by DamnDelicious'
```

Demonstration

Conclusion & Future Work

- * TF-IDF + LSA gives sensible recommendations

Conclusion & Future Work

- * TF-IDF + LSA gives sensible recommendations
- * Build mobile-friendly Flask web application

Conclusion & Future Work

- * TF-IDF + LSA gives sensible recommendations
- * Build mobile-friendly Flask web application
 - * Search for recipes by ingredients and cuisine type

Conclusion & Future Work

- * TF-IDF + LSA gives sensible recommendations
- * Build mobile-friendly Flask web application
 - * Search for recipes by ingredients and cuisine type
- * Scrape data from different sources and in different languages

Conclusion & Future Work



Thank you!



Bon Appétit!

Thank you!



Business case

Hellofresh

- * The world's leading meal kit provider
- * **65.6 million** meals in Q1 2019
- * 11 countries across 3 continents



Appendix - Motivation

“Gotchas”, issues worth noting, materials for blog:

- * Dimensionality Reduction
- * Before LSA and NMF, the dimension of the vectorized documents depends on how many unique word tokens there are
- * After dimensionality reduction, the corpus is transformed into a sparse matrix, where the number of features is the number of topics
- * A more rigorous way to determine the optimal number of topics is to use a method such as GridSearch
- * The top 5 topics generated at $k=10$ are almost identical to the ones of $k=5$, so I decided not to spend time tuning that parameter

Appendix

- * Cosine Similarity
- * This sklearn function can take either ndarrays or sparse arrays as inputs
- * Initially it was running really slowly for me, and crashed. This was because I only used one input, so it was calculating the pairwise cosine similarity between every single document inside the dataset and every other document. For our purposes, one-vs-all was sufficient (the similarity between the one selected recipe and every other recipe in the dataset)

Appendix

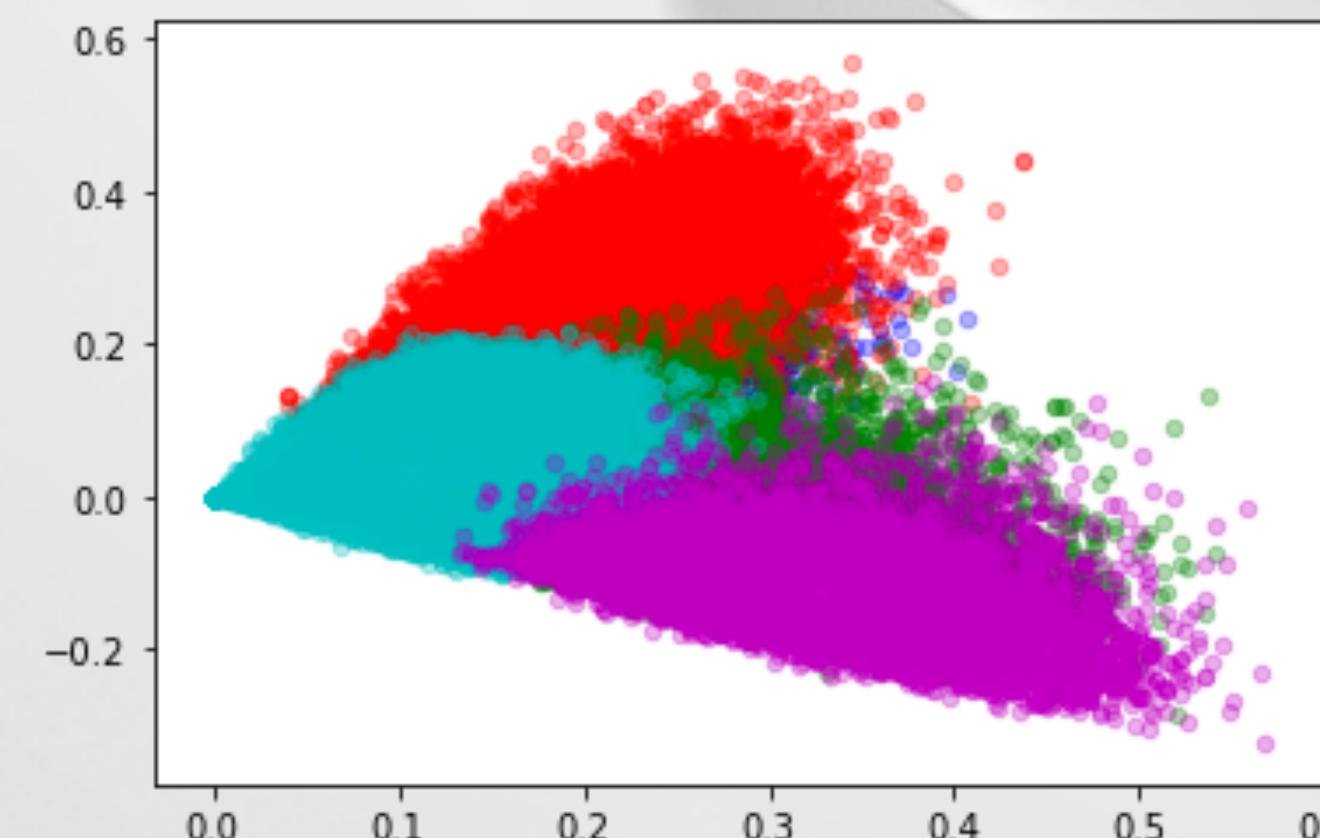
- * Soft Cosine Similarity
- * There is an option in Gensim to calculate soft cosine similarity, where words that have related meanings are also considered “equal” — “Hi” and “Hello”, “President” and “Prime Minister” are vectors in the same direction, with cosine similarity equal to one
- * But due to the nature of the dataset, this could lead to unique ingredients being lost in the comparison. So I decided not to implement it

Appendix

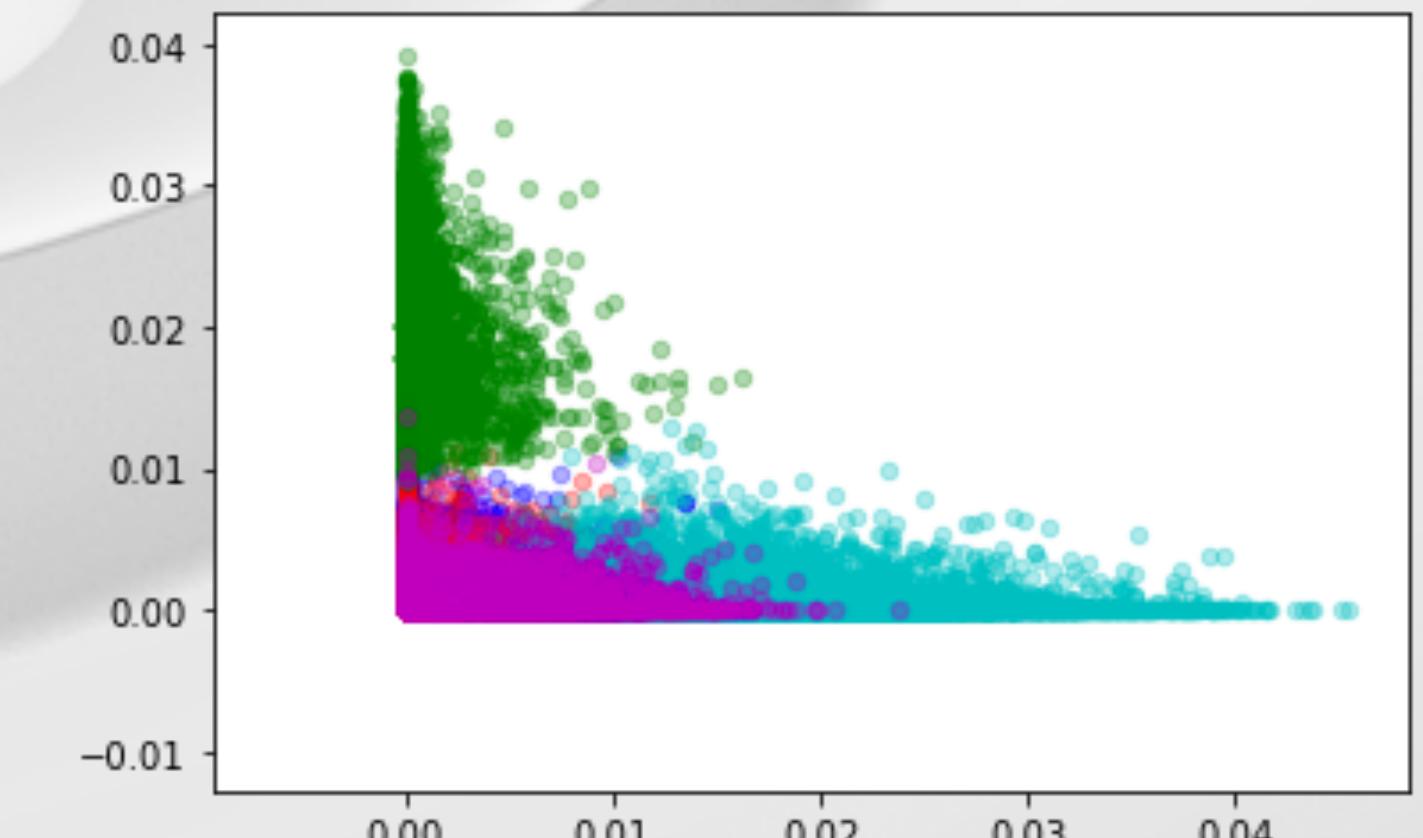
* Kmeans Clustering

- * I performed Kmeans clustering using both LSA and NMF, but decided to remove them from the submitted version of my code, as well as this presentation, because they do not help with my narrative
- * There remains the option to cluster the topics, then only find similar recipes (recommendations) within the cluster of the target recipe. I did not follow this path during the allotted time because my initial motivation was to recommend recipes with similar ingredients but from cuisine types other than that of the target. Also clustering sacrifices interpretability when visualized in high-dimensional space

LSA, k=5



NMF, k=5



Appendix

- * Cuisine Type Classifier
- * In my proposal and the EDA stage of this project, I mentioned building and training a classifier for cuisine types using the Kaggle What's Cooking dataset
- * This is because I wanted to use metrics from supervised learning after generating recommendations in unsupervised learning
- * It might be beneficial to use the cuisine type classifier to check the clusters, if the clusters are expected to be based on cuisine types

Appendix