

Cuisine Exploration Summary

Team Explorer

Alejandra Castillo, Anya Michaelsen,
Benjamin Sheller, Karan Srivastava

Motivation/Overview

There are increasingly more platforms today that make food-based predictions for their users, from restaurant recommendation systems that are based on users' cuisine preferences to platforms that identify potential cuisine options based on the ingredients in your refrigerator. Our team studied the relationship between the ingredients in a recipe and the recipe's cuisine in order to improve existing systems and explore such culinary relationships. The dataset we used was provided by Yummly for the Kaggle competition "What's cooking?", which contained almost 40,000 recipes from 20 cuisines using over 6,000 unique ingredients.

Classification

The primary goal of this project was to predict a cuisine using the ingredients in a recipe. Using recipes as sample points and cuisines as classes, we considered various classification models. The logistic regression model provided an accuracy score of 77.9%. Using the logistic regression coefficients, we also extracted top ingredients for each cuisine. For example, "coconut milk" and "fish sauce" were among the top 10 Thai ingredients. The Linear Support Vector Classifier provided an accuracy score of 76.8%. Other models considered were k -Nearest Neighbors, Random Forest, Linear and Quadratic Discriminant Analysis.

Clustering

Our second goal was to identify cuisine similarities with clustering algorithms. For this, we aggregated ingredient counts across all recipes within a cuisine, and encoded the data using term frequency - inverse document frequency (tf-idf) statistics, which weight values by how unique an ingredient is to a cuisine. We then applied k -Means clustering algorithms to group together similar cuisines, creating visualizations using principal component analysis for the dimensionality reduction. We further used hierarchical clustering to analyze how close or far multiple clusters were. For example, we saw that the cluster of Thai and Vietnamese were the closest cuisines within the Asian cluster which also included Korean, Japanese, and Chinese.

Applications

The ability to predict a cuisine using data from a recipe can be applied in several settings. Online recipe repositories such as, allrecipes.com, could implement one of our models to automatically create tags for recipes. In this way users can access all recipes with a particular cuisine tag. Applications and platforms that implement food or restaurant recommendations can use our models to identify the type of cuisine a user is interested in based on items on their grocery list or search history. For example, Yelp users who frequent Thai restaurants might be recommended Vietnamese places, and someone shopping for feta and black olives online might be recommended other top Greek cuisine ingredients, such as chickpeas and tahini.