CGAS MINI PROJECT

Food Pairing

Analyzing Ingredient Combinations for Optimal Food Pairings

11 December, 2024

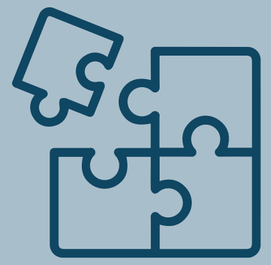
Introduction

This project uses NLP and Machine Learning to analyze ingredient combinations in thousands of popular recipes scraped from the Internet and identify the most common food pairings. The project visualizes food pairings and clusters by extracting ingredients and computing frequent combinations, providing data-driven insights into which ingredients go well together.

Project Deliverables

Ingredient Pairing Dataset

A curated dataset of popular ingredient combinations extracted from thousands of recipes.



Food Pairing Visualization

Interactive or static visualizations of common ingredient pairings and clusters, showing relationship between ingredientd .

Pairing Recommendation Tool

An ML/NLP-based tool that suggests ingredient pairings based on input ingredients.

Website

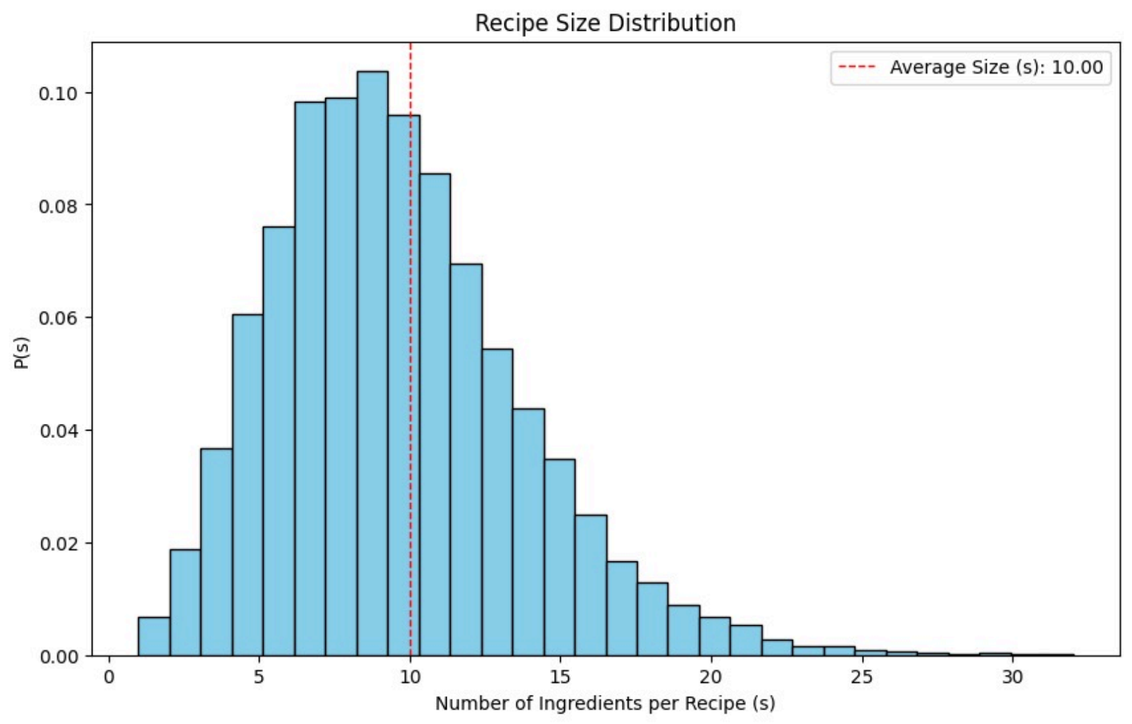
Users can input ingredients and receive pairing suggestions, along with visualized food pairing data.

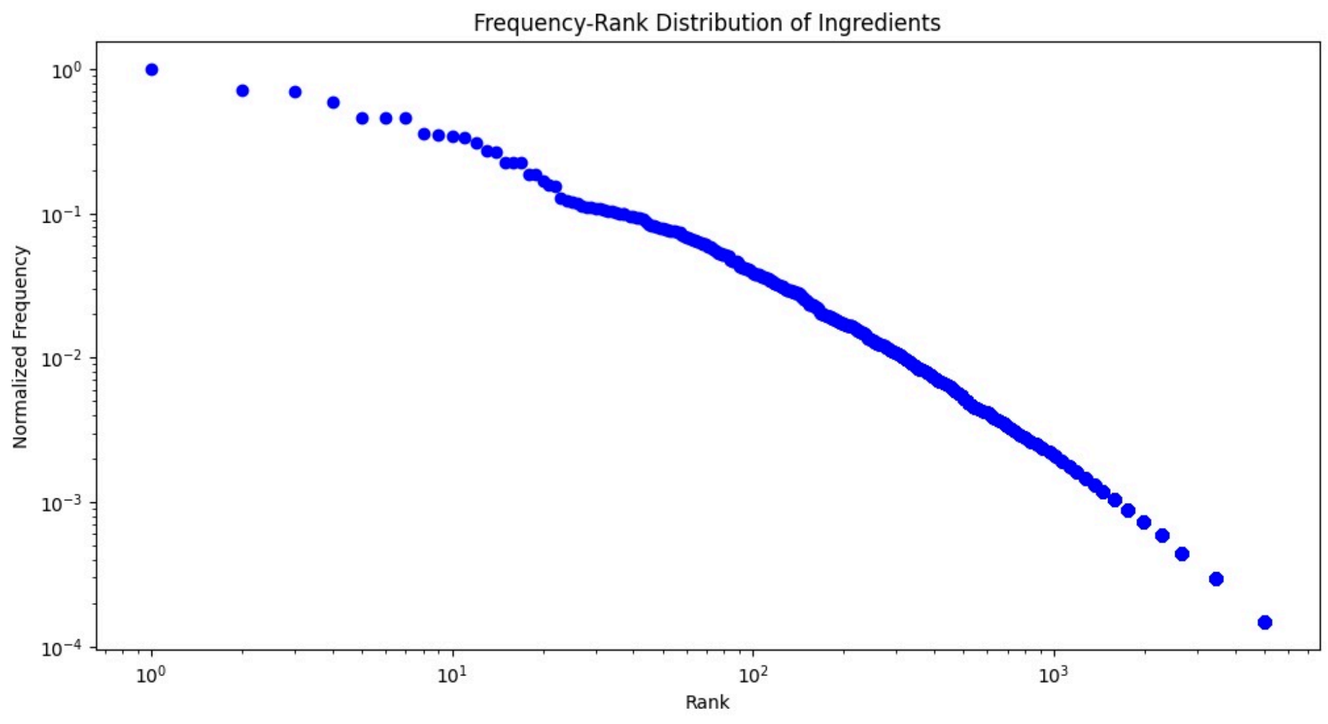
Dataset

Current scenario:

The dataset curation process begins by scraping recipe from AllRecipes using BeautifulSoup and regex. It then iterates through each category, retrieves recipe URLs from individual category pages, and extracts structured recipe data (such as headline, ingredients, instructions, and prep time) from the recipe pages using BeautifulSoup and JSON parsing. The data is formatted and written to a CSV file, ensuring consistency and organization. To handle potential connection issues, the script rotates through proxies and employs retry logic for HTTP errors, ensuring robustness. This curated dataset is then saved in a CSV file for further analysis or processing.

Total rows in dataset - 34,000





Recipe size distribution - The average recipe size is 10

Frequency-rank distribution graph follows the power law

Models

Apriori

Identifies frequent ingredient combinations by discovering association rules and analyzing co-occurrence patterns

KNN

Finds ingredient pairings by identifying similar recipes or ingredient clusters in a high-dimensional space.

DBSCAN

Clusters ingredients based on density, helping to uncover non-linear and non-uniform ingredient pairing patterns

Cosine Similarity

Measures the similarity between ingredient pairings based on their vector representation in the recipe dataset

Autoencoders

Reduces dimensionality to capture latent features of ingredient pairings, enabling discovery of complex relationships

Methodology

Data Collection:

- We gather recipe data, focusing on ingredient lists, to analyze common ingredient combinations across various dishes.

Analyze ingredient combinations identify the most common food pairings

- Iidentify the most frequently paired ingredients using techniques like association rule mining.

Visualizing food pairings

- Ingredient pairings are visualized through graphs, heatmaps, highlighting key connections and patterns in food combinations

Pairing Recommendation Tool

- Ingredient pairings are visualized through graphs, heatmaps, highlighting key connections and patterns in food combinations

Website

- The tool is hosted on a website where users can input ingredients and receive pairing suggestions, along with visualized food pairing data.

Visualisation

Ingredient Network Graphs:

- It is a co-occurrence graph, that visualizes the relationships and pairings among the top n most common ingredients based on their frequency of being mentioned together in recipes.

Heatmap of Pairing Frequencies:

- This heatmap visualizes the co-occurrence frequencies of the top n ingredients in recipes, highlighting which ingredients are most commonly used together.

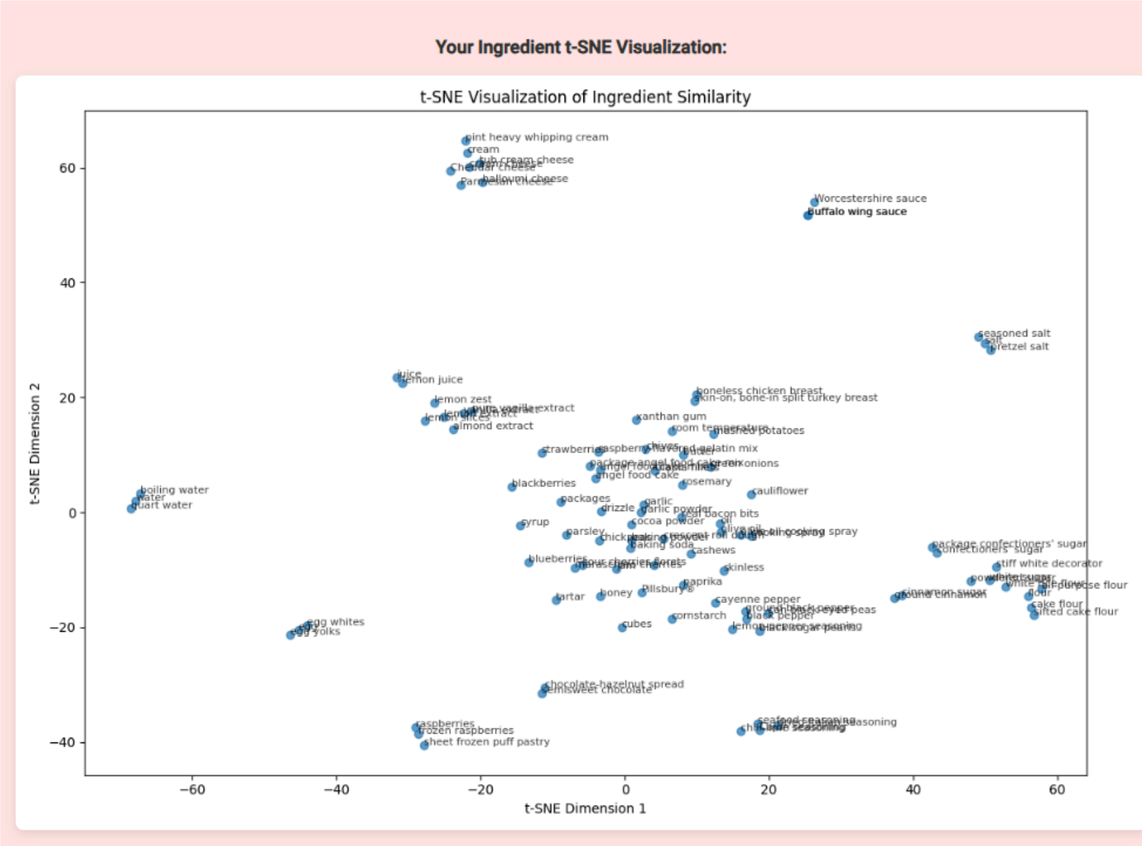
Word Clouds:

- This illustrates the frequency of ingredients in recipes, where the size of each word reflects how often the ingredient appears, emphasizing the most commonly used ingredients like “salt,” “pepper,” and “sugar.”

Ingredient Substitution

TSNE:

- As bigger the N goes, the clusters tsne forms in the data can help us find the substitutes for the ingredients



website

Conclusion

The project successfully leverages NLP and Machine Learning to analyze ingredient combinations and uncover optimal food pairings. By employing models such as Apriori, Cosine Similarity, KNN, Autoencoders, and DBSCAN, it identifies patterns and relationships within a curated dataset of recipes. The visualization tools and recommendation system provide actionable insights, enhancing the culinary creativity of users. With an interactive website showcasing pairing suggestions and visual data, the project demonstrates the potential of data-driven approaches in the culinary domain, offering valuable tools for chefs, food enthusiasts, and researchers alike.

Thank you