

Anamika Lal

BrainStation Capstone Sprint 2

# Recipe Recommender

# General Overview

- Build a recipe recommendation system such that it can:
  - Help reduce food waste
  - Promote mindful / conscious eating
  - Encourage users to try new recipes

# Proposed Vision using ML

- “Reduce food waste” –
  - Use existing ingredients.
  - ‘Rank’ recipes that have less ingredients higher.

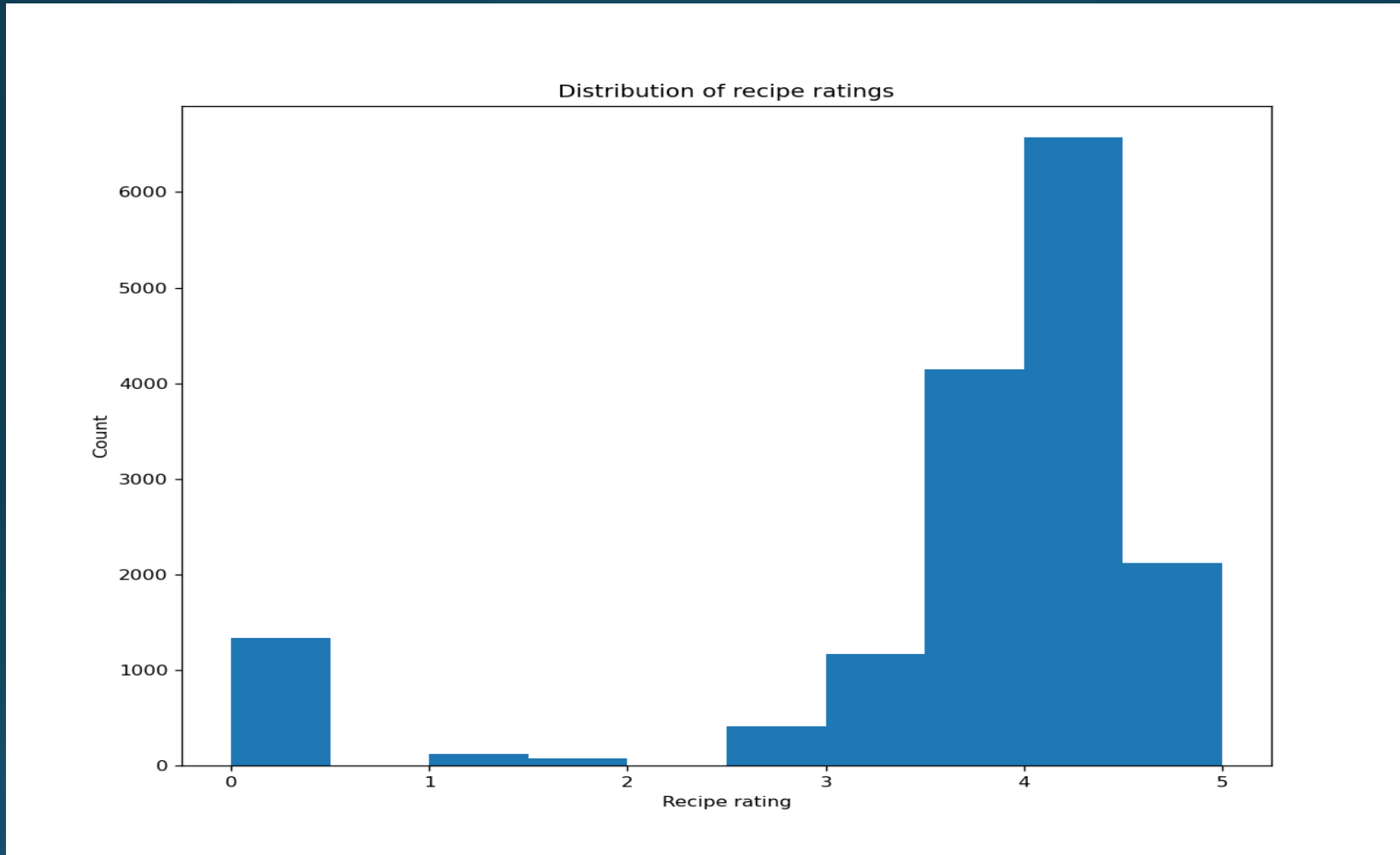
# Overview of Dataset

- Epicurious – Recipes with Rating and Nutrition (Kaggle)
  - Full text for steps and ingredients with measurements.
  - 14K recipes.

```
df.head()
```

	recipeId	calories	rating	title	directionsStr	categoriesStr	ingredientsStr
0	0	426.0	2.500	Lentil, Apple, and Turkey Wrap	['1. Place the stock, lentils, celery, carrot,...	['Sandwich', 'Bean', 'Fruit', 'Tomato', 'turke...	['4 cups low-sodium vegetable or chicken stock...
1	1	403.0	4.375	Boudin Blanc Terrine with Red Onion Confit	['Combine first 9 ingredients in heavy medium ...	['Food Processor', 'Onion', 'Pork', 'Bake', 'B...	['1 1/2 cups whipping cream', '2 medium onions...
2	2	165.0	3.750	Potato and Fennel Soup Hodge	['In a large heavy saucepan cook diced fennel ...	['Soup/Stew', 'Dairy', 'Potato', 'Vegetable', ...	['1 fennel bulb (sometimes called anise), stal...
3	4	547.0	3.125	Spinach Noodle Casserole	['Preheat oven to 350°F. Lightly grease 8x8x2-...	['Cheese', 'Dairy', 'Pasta', 'Vegetable', 'Sid...	['1 12-ounce package frozen spinach soufflé, t...
4	5	948.0	4.375	The Best Blts	['Mix basil, mayonnaise and butter in process...	['Sandwich', 'Food Processor', 'Tomato', 'Kid-...	['2 1/2 cups (lightly packed) fresh basil leav...

# 54% of dataset has ratings $> 4.0$



# Overview of Preprocessing

- Recipes in JSON, imported to Pandas
- Convert lists to strings
- Choosing a Vectorizer
- Custom Tokenization

# Baseline models and evaluation metrics

- Unsupervised problem, no target variable
- Evaluation of result quality done manually



# Baseline models and evaluation metrics

- Given a recipe name, can we find the 10 most similar recipes from the dataset?
- Cosine distance
- NearestNeighbors

Enter the recipe name

Chicken Parmesan

Chicken Parmesan : 0.0 : 5.0

Quick Baked Chicken Parmesan : 0.423 : 5.0

Fried Chicken Biscuits : 0.438 : 5.0

Fried Chicken Thighs with Cheesy Grits : 0.464 : 0.0

Chicken and Dumplings with Mushrooms : 0.47 : 4.375

Spanish-Style Fried Chicken With Grilled Avocado : 0.481 : 5.0

3-Ingredient Shakshuka : 0.481 : 3.125

Fusilli with Shrimp and Paneed Chicken : 0.492 : 4.375

Horseradish-Yogurt Sauce : 0.497 : 0.0

Lamb and Eggplant Casserole (Moussaka) : 0.504 : 3.75

Spicy Oven-Fried Chicken : 0.505 : 4.375

9,337	12,425	302	4.375	Cinnamon Biscuits	['Blend 2/3 cup sugar, flour, cake f
-------	--------	-----	-------	-------------------	--------------------------------------

## Using pretrained models with user input

Enter the ingredients to include below

Shrimp, Zucchini

Enter the ingredients to exclude below

Tomatoes

Shrimp and Mango Skewers with Guava-Lime Glaze : 0.677 : 4.375

Grilled Rosemary Garlic Shrimp : 0.728 : 4.375

Tagliatelle with Zucchini and Shrimp : 0.742 : 3.125

Zucchini and Parmesan Skillet Cake : 0.742 : 3.125

Carrot and Squash Ribbons : 0.742 : 3.125

Shrimp Salad with Zucchini and Basil : 0.748 : 4.375

Shrimp Olivier : 0.762 : 3.75

Chicken Breasts with Rock-Shrimp Sauce : 0.764 : 3.75

Cream of Zucchini and Anise Soup : 0.764 : 3.125

Pasta with Shrimp and Basil Vinaigrette : 0.774 : 4.375

Hot Cauliflower with Shrimp : 0.774 : 3.125

# Next Steps

- Advanced Modeling
  - Penalize recipes based on various criteria
  - Tokenization to handle plurals
  - Use categories information also
  - Improve Streamlit app

Demo