SmartRecipes

Capstone Sprint 3

Anamika Lal

General Overview

Build a recipe recommendation system such that it can help reduce food waste.

Solution using ML

- Chose a dataset with ~ 20,000 recipes.
- Built a content based recommendation system.
  - Custom tokenizer for CountVectorizer.
  - Sklearn NearestNeighbors learner with cosine metric
- The recommender gives recipes from the dataset, ranked by closest match and less ingredients.

Modeling and evaluation

- Unsupervised problem, no target variable.
- Evaluation of result quality done manually
- Recipes with least ingredients show up in the top 5 results ~ 80% of the time.

Demo

Challenges

1. Selecting a dataset – Looked at 6 different options, did EDA on three.

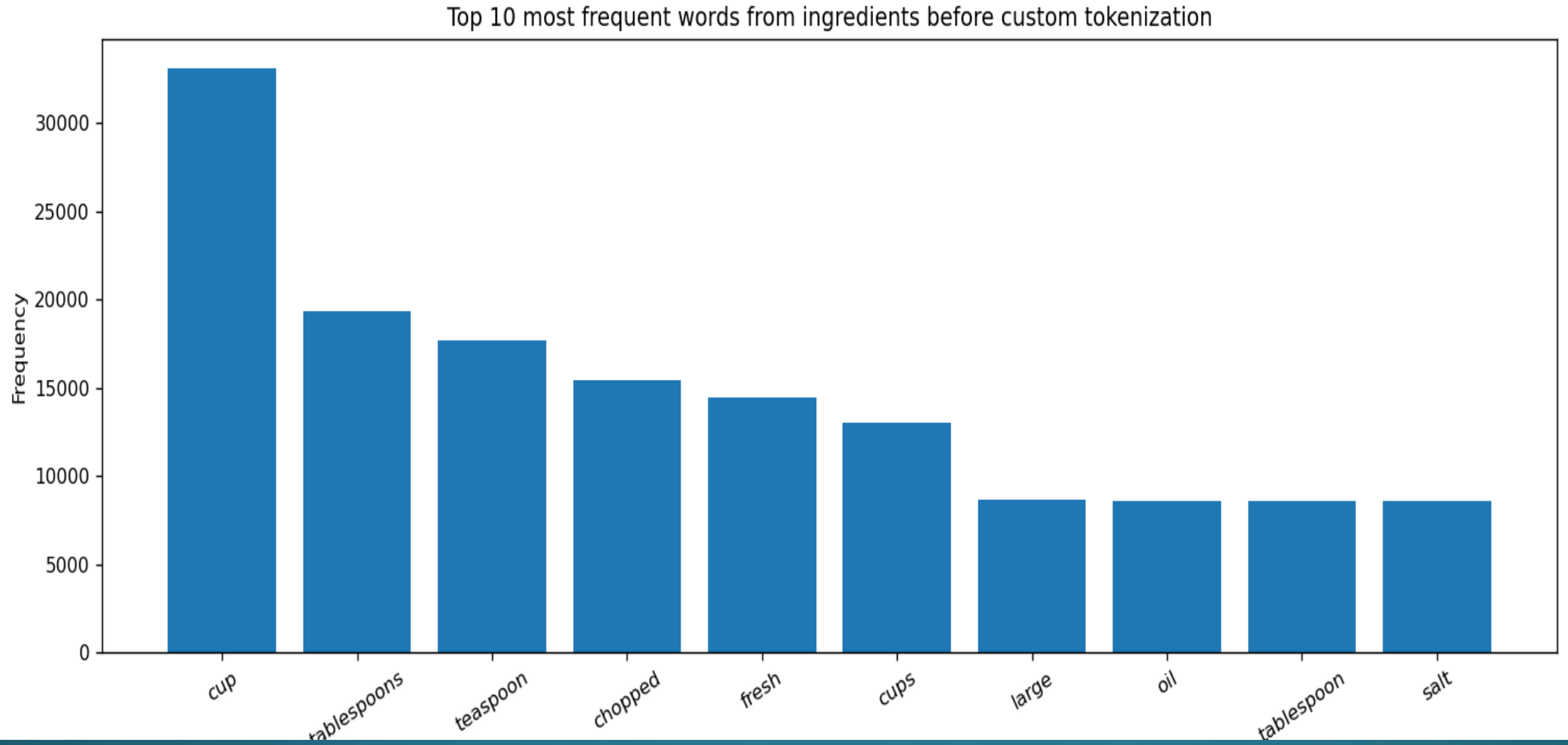
- Full text for steps of the recipe
- Ingredients should have measurements in them
- Some category information should also be there e.g. gluten-free

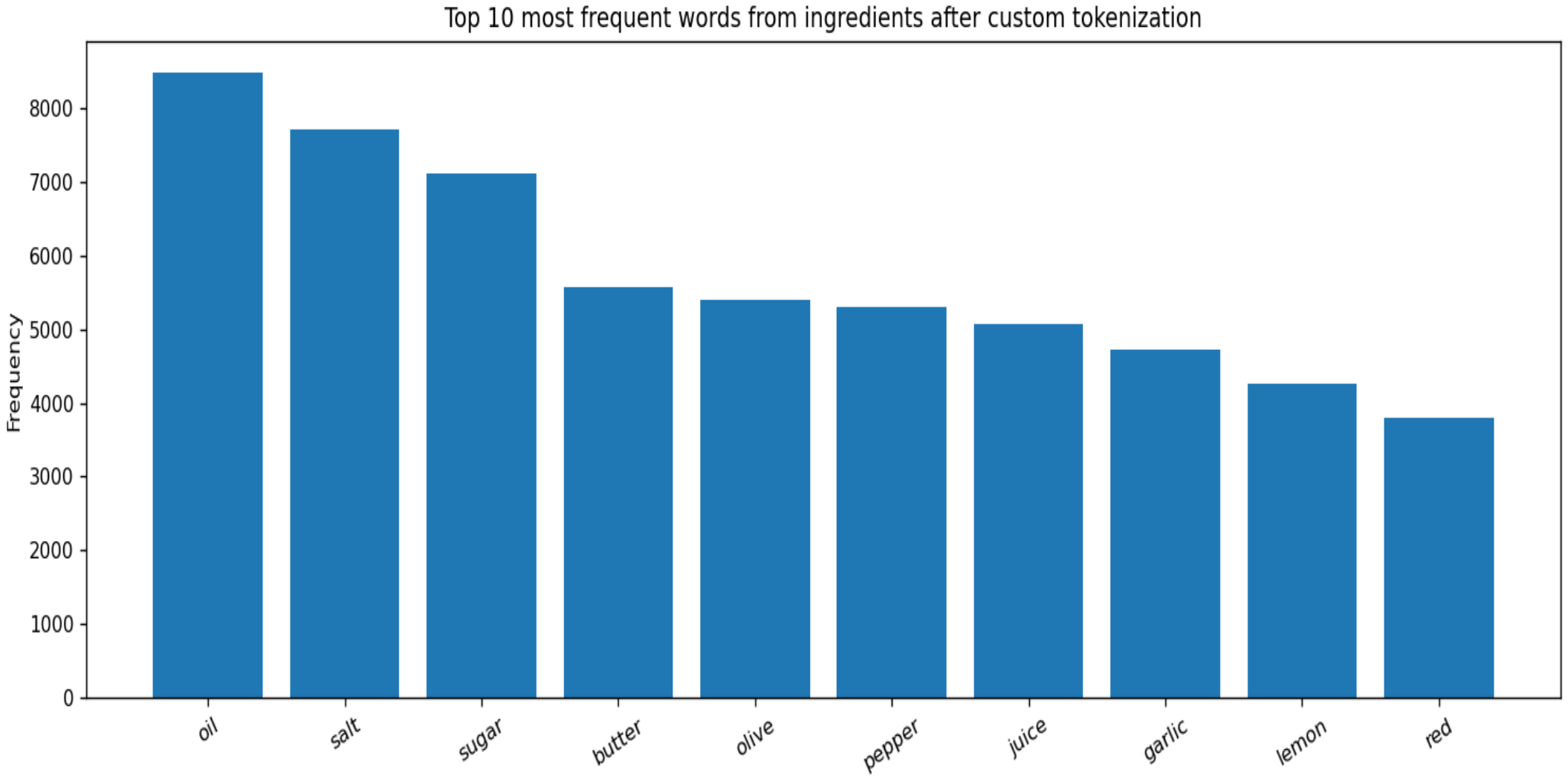
Challenges

2. Custom tokenization – had to iterate many times

- Text about measurements was messing up the results.

- Since the ingredients were in full text, there are many action words like ‘chopped‘, `peeled` etc.


Top 10 most frequent words from ingredients before custom tokenization
Frequency
30000
25000
20000
15000
10000
5000
0
cup
tablespoons
teaspoon
chopped
fresh
cups
large
oil
tablespoon
salt


Top 10 most frequent words from ingredients after custom tokenization
Frequency
0
1000
2000
3000
4000
5000
6000
7000
8000
oil
salt
sugar
butter
olive
pepper
juice
garlic
lemon
red

Challenges

3. Handling two word ingredients and categories

   - Two word ingredients e.g. ‘peanut butter’, ‘bell pepper’ (Bigrams did not help)
   - Lots of irrelevant categories in the dataset.
   - Solved using custom vocabulary.

Challenges with user text input

- Implemented spell checker using jaccard distance.

- Handled plurals using the `pattern` package.

Conclusion

- Built a Streamlit app for the recommendation system.
- Recommender returns 10 recipes.
- Recipes with least ingredients show up in the top 5 results ~ 80% of the time.

Thank you !