

CuisineHelper

“Re-discover your pantry”

Uygar Sozer

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

Motivation

- Target audience:
 - Young professionals, mid-to-late 20s
 - Home cooks with a wide palate
 - Anyone who's tired of eating the same thing!
- People can match a short list of ingredients with other ingredients
 - Could be other common ingredients
 - Hopefully also some cuisine-specific ones
- "Predicts" the cuisine in mind as well
 - Recommendation at the forefront of the prediction

Link to app

<http://cuisi-publi-1adtzdx2y8p7v-963265392.us-east-1.elb.amazonaws.com/>

Data

```
processing:
  clean:
    PATTERNS:
      - '^(\.?*\ )'
      - ',.*$'
    REMOVE_WORDS:
      - 'low-fat'
      - 'lowfat'
      - 'low fat'
      - 'sodium'
      - 'lb\.'
      - 'ounc'
      - 'fresh'
```

- Dataset from Kaggle
- Contains entries for each recipe, each entry consists of ingredients and a cuisine label

```
{'id': 10259, 'cuisine': 'greek', 'ingredients': ['romaine lettuce', 'black olives', 'grape tomatoes', 'garlic', 'pepper', 'purple onion', 'seasoning', 'garbanzo beans', 'feta cheese crumbles']}
```

- Around 33k recipes, could be extended
- Cleaning:
 - regex matching
 - remove “common” words

The model

	ingredient	cuisine
0	romaine lettuce	greek
1	black olives	greek
2	grape tomatoes	greek
3	garlic	greek
4	pepper	greek
...
428270	garlic	mexican
428271	white sugar	mexican
428272	roma tomatoes	mexican
428273	celery	mexican
428274	dried oregano	mexican

428275 rows × 2 columns



	cuisine	brazilian	british	cajun_creole	chinese	filipino	french	greek	indian	irish	italian
ingredient											
Gochujang base		0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
Italian bread		0.0	0.0	9.0	0.0	0.0	6.0	1.0	0.0	1.0	89.0
Italian parsley leaves		2.0	1.0	3.0	1.0	1.0	17.0	1.0	1.0	1.0	74.0
Mexican cheese blend		0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mexican oregano		0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0

- Recipes (ingredients & labels) converted to a single table
- Count the occurrence of ingredients within each cuisine, drop if total does not pass a threshold (100 in the final model)
- Table stored in DB
- Model mean centers each entry, and converts each column to equal to 1, then scale it up by a constant (1000 in the final model)
- **Predictions:** When ingredients are entered, simply add up all the “cuisine scores”, highest 3 is selected
- **Recommendations:** 5 top ingredients within cuisine that doesn’t include what’s already selected

Insights

```
processing:
  clean:
    PATTERNS:
      - '^(\..*?\)' '
      - ',.*$'
    REMOVE_WORDS:
      - 'low-fat'
      - 'lowfat'
      - 'low fat'
      - 'sodium'
      - 'lb\.'
      - 'ounc'
      - 'fresh'
  features:
    DROP_ROWS:
      - 'salt'
      - 'water'
      - 'sugar'
      - 'olive oil'
      - 'garlic cloves'
    MIN_PREVALENCE: 100
```

- Model achieves 70% success if all ingredients are included
 - Success: the label was correctly predicted from each recipe (including all the ingredients) within the top 3 choices
- However, some very common choices will then shoot up to top for each cuisine
- Better to take out some words. In the configuration shown here, 68% success
- Generally does a good job of predicting relevant cuisines and give accurate recommendations. Tested by amateur home cooks!

Thank you!

Uygar Sozer

[GitHub](#)



Uygar Sozer

usoz

Edit profile

5 followers · 6 following · 12

Northwestern University

Evanston, IL

Highlights

☆ PRO

Organizations

