What's Cooking

Jincai Ma

Xi'an Shiyou University

November 21, 2020

Introduction

Introduction
Data Import And Introduction
Analyze Data
Data Visualization
Build Model

- Use recipe ingredients to categorize the cuisine.
  Given the name of the condiment, predict the cuisine to which the dish belongs.
- In the dataset, including the recipe ID, the dish, and the list of ingredients for each recipe (variable length).The data is stored in JSON format.
  1.train.json- A training set that contains the recipe ID, dish type, and ingredient list
  2.test.json- A test set containing a recipe ID and a list of ingredients
  3.sample_submission.csv-Properly formatted sample submission document

Data Import And Introduction

Introduction
Data Import And Introduction
Analyze Data
Data Visualization
Build Model

- Import the JSON file with Pandas: We can get the data set of dish names, including 39774 training data and 9944 test samples.
- To see the distribution of our data set and the total variety of dishes, we printed out some of the data samples.

```
      id      cuisine  \
0  10259        greek
1  25693  southern_us
2  20130     filipino
3  22213       indian
4  13162       indian

                                                                                                ingredients
0  [romaine lettuce, black olives, grape tomatoes, garlic, pepper, purple onion, seasoning, garbanzo beans, feta cheese...
1  [plain flour, ground pepper, salt, tomatoes, ground black pepper, thyme, eggs, green tomatoes, yellow corn meal, mil...
2  [eggs, pepper, salt, mayonaise, cooking oil, green chilies, grilled chicken breasts, garlic powder, yellow onion, so...
3                                                                                [water, vegetable oil, wheat, salt]
4  [black pepper, shallots, cornflour, cayenne pepper, onions, garlic paste, milk, butter, salt, lemon juice, water, ch...
```

- Total dish classification

  There are 20 dishes in total, which are: [’brazilian’ ’british’ ’cajun_creole’ ’chinese’ ’filipino’ ’french’ ’greek’ ’indian’ ’irish’ ’italian’ ’jamaican’ ’japanese’ ’korean’ ’mexican’ ’moroccan’ ’russian’ ’southern_us’ ’spanish’ ’thai’ ’vietnamese’]

Analyze Data

Introduction
Data Import And Introduction
Analyze Data
Data Visualization
Build Model

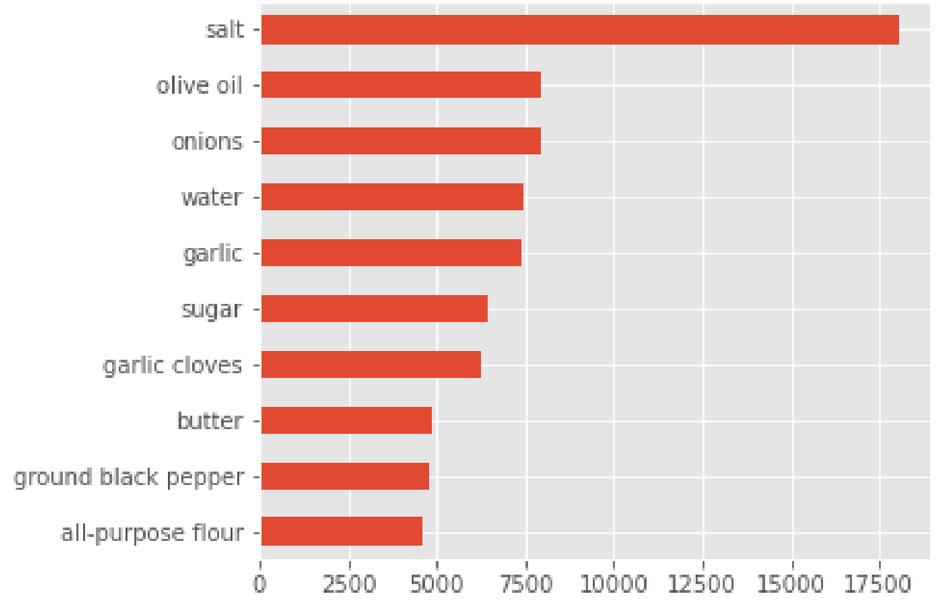
- The data set is divided into Features and Target Variables.
- Features:'ingredients', we were given the names of the ingredients contained in each dish. Target variable:'cuisine', is the classification of cuisines that we want to predict.
- Extract the Feature of training data set into train_integredients variable Extract the Target Variables into the train_Targets variable

```
0       [romaine lettuce, black olives, grape tomatoes, garlic, pepper, purple onion, seasoning, garbanzo beans, feta cheese...
1       [plain flour, ground pepper, salt, tomatoes, ground black pepper, thyme, eggs, green tomatoes, yellow corn meal, mil...
2       [eggs, pepper, salt, mayonaise, cooking oil, green chilies, grilled chicken breasts, garlic powder, yellow onion, so...
3                                                                                                  [water, vegetable oil, wheat, salt]
4       [black pepper, shallots, cornflour, cayenne pepper, onions, garlic paste, milk, butter, salt, lemon juice, water, ch...
                                                                                                  ...
39769   [light brown sugar, granulated sugar, butter, warm water, large eggs, all-purpose flour, whole wheat flour, cooking ...
39770   [KRAFT Zesty Italian Dressing, purple onion, broccoli florets, rotini, pitted black olives, Kraft Grated Parmesan Ch...
39771   [eggs, citrus fruit, raisins, sourdough starter, flour, hot tea, sugar, ground nutmeg, salt, ground cinnamon, milk, ...
39772   [boneless chicken skinless thigh, minced garlic, steamed white rice, baking powder, corn starch, dark soy sauce, kos...
39773   [green chile, jalapeno chilies, onions, ground black pepper, salt, chopped cilantro fresh, green bell pepper, garlic...
Name: ingredients, Length: 39774, dtype: object
0             greek
1       southern_us
2          filipino
3            indian
4            indian
           ...
39769         irish
39770       italian
39771         irish
39772       chinese
39773       mexican
Name: cuisine, Length: 39774, dtype: object
```

Data Visualization

Introduction
Data Import And Introduction
Analyze Data
Data Visualization
Build Model

- What are the top 10 most frequently used seasonings.



Build Model

Introduction
Data Import And Introduction
Analyze Data
Data Visualization
Build Model