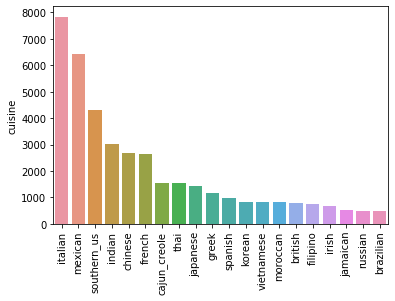
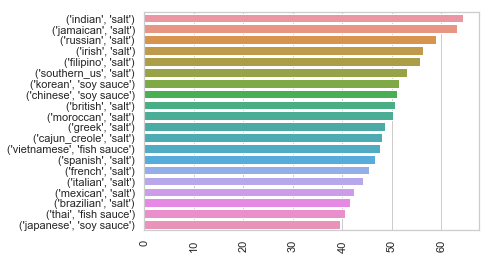
What’s Cooking?

Summary Sheet

**Objective:**

Ever wonder what type of cuisine you’re currently eating? The objective of this study was to predict the type of cuisine based on the ingredients in the recipe.

The Data: 

The data set is a json file containing the recipe id, type of cuisine and the list of ingredients. Most of the recipes were classified as Italian (7838 recipes), Mexican (6438) and Southern US (4320) contrary to Russian (489) and Brazilian (467). The top 5 ingredients were salt, water, onions, olive oil and garlic. 

Salt being the most common ingredient, we can see it was most commonly used in 64.4% of Indian cuisine and the least in Japanese.

Data Processing:

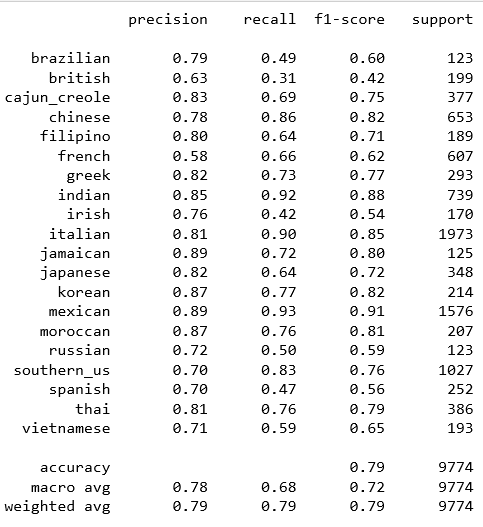
1. Special characters, numbers, capital letters
2. Stop words: removal of descriptive words.
3. Stemming: removal of singular, plural, past and present tense.

Data was then separated into training and testing 

1. Encoding cuisines and recipes:  Cuisines were encoded in the simplest way, i.e. as integers. Each zero or one stands for whether an ingredient was or was not used in the recipe.

Logistic Regression:

We have a multiclass classification problem of predicting one of 20 cuisines by the recipe. The logistic regression analysis is used for classification problems with binary dependent variables.  To predict the likelihood of the recipe belonging to a specific cuisine, we created 20 logistic regression models, one for each cuisine. All these 20 models were created using the Logistic Regression object from sklearn package. The below figure represents the performance of our model on the test dataset.

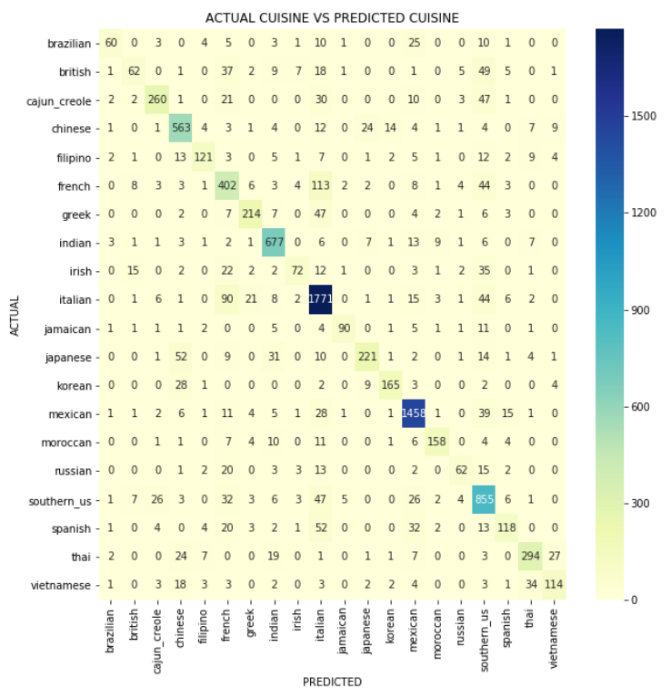


The overall accuracy of our model is 79%, which appears to be a good result given that we have 20 cuisines.

Highest precision and recall for Indian, Italian and Mexican cuisines, due to large sample size

The below confusion matrix represents how cuisines were predicted by our model. On the vertical axis there are actual cuisines (labels), while on a horizontal axis there are predicted cuisines.

French cuisine was falsely predicted as Italian 113 times while Italian cuisine was falsely predicted as French 90 times.



Conclusion:

In the end we were successfully able to create a model with approximately 79% accuracy of determining a type of cuisine based on ingredients.  With additional recipes the model will only improve and yield the intended results. Never forget a dish again!