Cuisine Classification by Ingredients

Problem: Given a list of recipe ingredients, determine the cuisine they are from.

Approach: We trained multiple classifiers on our data: decision trees (and related ensemble methods), Naive Bayes, logistic regression, and linear SVM.

We built a heat map in order to visualize the most likely regions of origination.

Heat map visualization of an Italian recipe (not including Mexico).

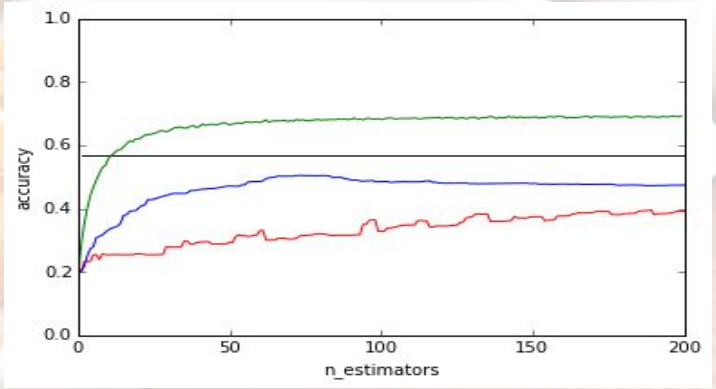
Results and Model Performance

Various versions of word lemmatization were used on the ingredients before running algorithms.

Lemmatization Version	*Logistic Regression Accuracy*	*Naive Bayes Accuracy*	*Linear SVM Accuracy*
1	0.777	0.715	**0.778**
2	0.774	0.739	**0.774**
None	0.729	0.731	**0.772**

The linear SVM was the most accurate of the classifiers we tested.

We built a learning curve to visualize optimal parameters for our decision tree and ensembles.



Green (Random Forest), Blue (AdaBoost-SAMME.R), Red (AdaBoost-SAMME), Black (Decision Tree)