

What is Cooking?

Team: Tyler Paulsen

tjp5115@rit.edu

“What is Cooking” will be creating a classification model to determine what ingredients go with a particular type of food -- Italian, Indian, etc. The project was taken from the website kaggle.com; a website used by data scientists to solve different types of data science problems. The challenge was created by a website called yummlly.com -- a food website that has a plethora of recipies.

Yummly provided two different datasets to Kaggle: a training set, and a test set. The training set is to train the model, and the test set is used to determine how well the model performed. Once a model has been created and been applied to the training set, it can sent to kaggle.com to determine the accuracy of the model.

The model will be determining what ingredients go to different types of recipes. The challenge to this will be creating a model that can deal with a large number of crossovers between ingredients and types of recipes. Since an ingredient is normally used between multiple recipes, it will be tough to find a grouping of ingredients that will used to predict a recipe. Overfitting will become something to be aware of because of the large amount of overlap between ingredients.

For a baseline, the submission of the data to Yummly will determine how accurate the model is. During the submission of the test data, Yummly will respond with an accuracy rating that will determine how accurate the model was at determining the type of recipe. So far, the top accuracy rating is: .81818.

I thought I heard during the class on Wednesday 9/23 that we could work alone if we preferred. I plan on working alone, unless another student has the same idea and is also working alone, then I will group up with him or her.

Link to project: <https://www.kaggle.com/c/whats-cooking>

About Yummly: <https://www.yummly.com/about/>