Sign up for a free account at Kaggle.com.

Next sign up for the "What's for Dinner?" competition at:

[https://www.kaggle.com/t/328e640dc717461fa620accca58c0c45 (Links to an external site.)Links to an external site.](https://www.kaggle.com/t/328e640dc717461fa620accca58c0c45)

Form a team of 1-3 people, and create a team on Kaggle and in this Canvas assignment.

For the first part of the competition, you are looking to create a submission that has at least 83% accuracy on the hidden test data using a k-Nearest Neighbor model.  Please see the discussion forum in the Kaggle contest for Python notebooks with code and output for:

1) Data exploration

2) Cross-validation for Model Tuning

3) Creation of a submission file

Here are the "knobs" in the preprocessing steps that you may want to experiment with during this first phase of the competition:

- Use of CountVectorizer vs. TfIdfVectorizer (as shown in the exploratory data analysis example code)

- Use of a preprocessor that stems words

- CountVectorizer parameters: ngram\_range, stop\_words, lowercase, analyzer, min\_df, max\_df, binary, max\_features.  See [http://scikit-learn.org/stable/modules/generated/sklearn.feature\_extraction.text.CountVectorizer.html (Links to an external site.)Links to an external site.](http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html) for details on these parameters.

- TfIdfVectorizer paramaters:  ngram\_range, stop\_words, lowercase, analyzer, min\_df, max\_df, binary, max\_features, use\_idf, smooth\_idf, sublinear\_tf.  See [http://scikit-learn.org/stable/modules/generated/sklearn.feature\_extraction.text.TfidfVectorizer.html (Links to an external site.)Links to an external site.](http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html) for details on these parameters.

Here are the "knobs" in the modelling steps that you may want to experiment with during this first phase of the competition:

- KNeighborsClassifier: n\_neighbors, weights, p.  See [http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier (Links to an external site.)Links to an external site.](http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier) for details on these parameters

 - Weights on attributes: To weight an attribute, you can multiply its value by a factor, e.g.:

ndx=v.get\_feature\_names().index('salt')

df\_x[:,ndx] = df\_x[:,ndx] \* 20

You can submit up to 2 submissions per day to Kaggle.  You do not want to rely on Kaggle for parameter selection and an estimate of error.

Make sure to select the submission that you want graded in Kaggle.

You should submit the following to this dropbox:

- Code used to select model parameters

- Code used for final model selection

- Your Kaggle team name (in the text entry box)