

Homework 3: Text Classification 1

Objectives:

- Gain experience with sklearn using text data
- Gain experience with text classification

Turn in:

- This program should be created in a notebook (Jupyter, Google, or Kaggle)
- Print to pdf and upload your pdf to eLearning

Instructions:

1. Go to Kaggle.com or any other resource. Find a text classification data set that interests you. I suggest a smaller data set. Divide into train/test. Create a graph showing the distribution of the target classes. Describe the data set and what the model should be able to predict.
2. Try Naïve Bayes, Logistic Regression, Neural Networks using sklearn.
3. In a text block in your notebook, write up your analysis of the performance of various approaches

Grading Rubric:

- Your grade is not determined by the accuracy achieved, but by how much work and thought you put into it

Caution: All course work is run through plagiarism detection software comparing students' work as well as work from previous semesters and other sources.