Final Project

Milestone 1

Group members: Kevin De Angeli and Hector D. Ortiz-Melendez

Data set: Sentimental Labelled Sentences Data Set.

The data set contains sentences labelled with positive or negative sentiment (0 or 1). This is an NLP classification problem. The data set is composed of reviews from Amazon, IMDB, and Yelp. We will use bag-of-words to extract features and run it through the different ML algorithms.

Link to the dataset: https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences#

Note: this is not one of the data sets suggested by Dr. Qi, but we are interested in NLP and this seems to be an interesting problem. If you think this data set is not appropriate, we can pick a different one.