

**Title:** Comparing Effectiveness of Different Classification Algorithms (Random Forests, SVM, Least Squares) with Different Feature Representation (Bag-of-words, TF-IDF) for Natural Language Processing.

**Final Project Group 22**

**Ahmad Zaidi,** [azaidi4@wisc.edu](mailto:azaidi4@wisc.edu), 907 472 3785

**Matthew Isaac,** [misaac3@wisc.edu](mailto:misaac3@wisc.edu), 907 471 7787

**Learning Objectives:**

- (1) Understanding Basics of NLP: Students will learn about NLP vocabulary (unigrams, tokenization, corpus), feature extraction from raw data (tweets) and stemming.
- (2) Understanding Random Forest Algorithm: Students will learn about decision trees and Random Forest Classification. They will compare the error rates between other classification algorithms such as least squares, SVM.
- (3) Understanding Different Feature Representations: Students will learn the shortcoming of Bag-of-words and how TF-IDF overcomes them for NLP.