Text Mining

Basic Idea of Project

- Use text mining / natural language processing to identify the author of specific blocks of text.
- Various techniques were used such as: Naive Bayes, Logistic Regression, SVM, SVM with Stochastic Gradient Descent, Random Forest, and Gradient Boosting, Bag-of-Words, Continuous Bag-of-Words (CBOW), TF-IDF, uni- and bi-grams.

Skills Learned

- Natural Language Processing (in Python)
- Regex and string processing