

CS 229 - Project Milestone

Bryan McCann
Stanford University
bmccann@stanford.edu

Brandon Ewonus
Stanford University
bewonus@stanford.edu

Nat Roth
Stanford University
nroth@stanford.edu

Abstract

In this report we analyze the political speeches made by members of the Democratic and Republican parties in the United States. Specifically, we attempt to learn which features best differentiate speeches made by the two parties, and develop a model to classify speeches as either Democrat or Republican.

1 Introduction

Division among the political parties in the United States has become an increasingly large problem.

The rhetoric...how partisan different congressmen are...etc.

2 Data Collection

where data came from? why we need a lot of data? how we're storing the data? stop words? stemming? etc.

3 Methods / Analysis

3.1 Naive Bayes

3.2 SVM

3.3 K-means

3.4 Logistic Regression

3.5 Regularization

4 Results / Discussion

5 To Do

References

- [1] "scikit-learn: Machine Learning in Python".
<<http://scikit-learn.org/stable/index.html>>
- [2] "History & Politics Out Loud: Famous Speeches".
<<http://www.wyzant.com/resources/lessons/history/hpol/>>

- [3] "American Rhetoric Speech Bank".
<<http://www.americanrhetoric.com/>>
- [4] "Presidential Rhetoric".
<<http://www.presidentialrhetoric.com>>
- [5] "The American Presidency Project".
<<http://www.presidency.ucsb.edu/index.php#axzz2i2nXPc43>>