Classifying Game of Thrones Subreddits

Gwen Rathgeber - The Matt

Instructor: Mahdi Shadkam-Farrokhi
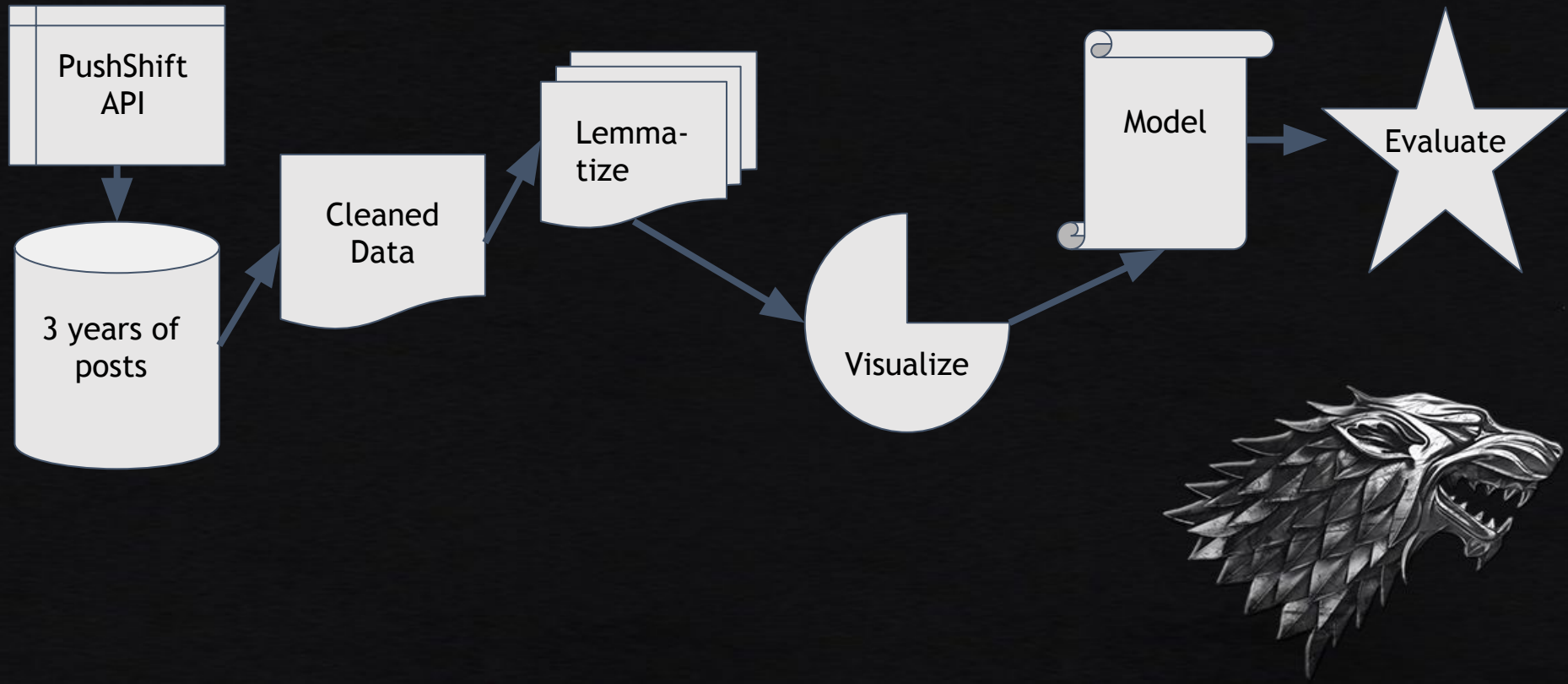
Problem Statement

- Reddit wants to remove redundant subreddits to improve user experience

- r/gameofthrones, r/freefolk, and r/asoiaf moderators insist that they are different spaces despite their shared topic

- To test this claim, we will see if we can predict the subreddit of a post
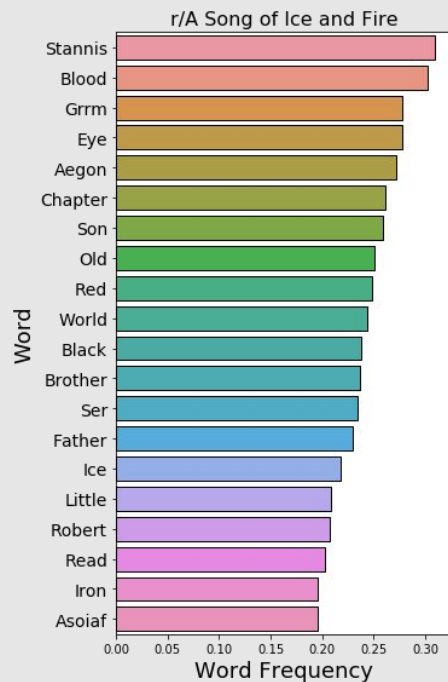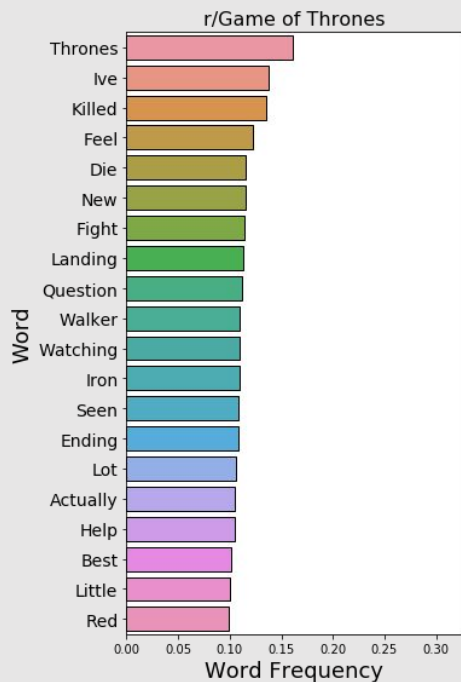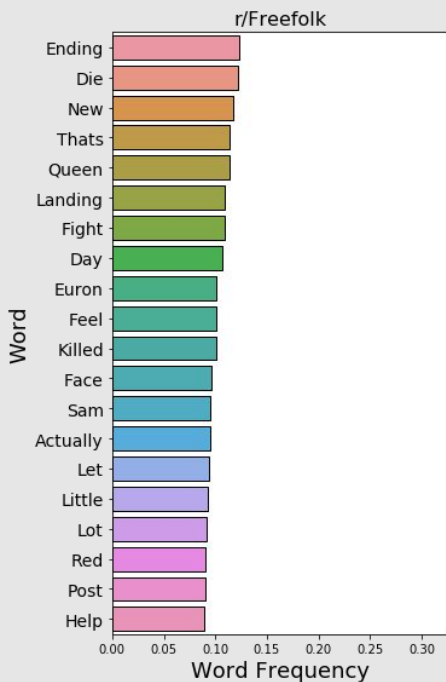
Workflow

PushShift API

3 years of posts

Cleaned Data
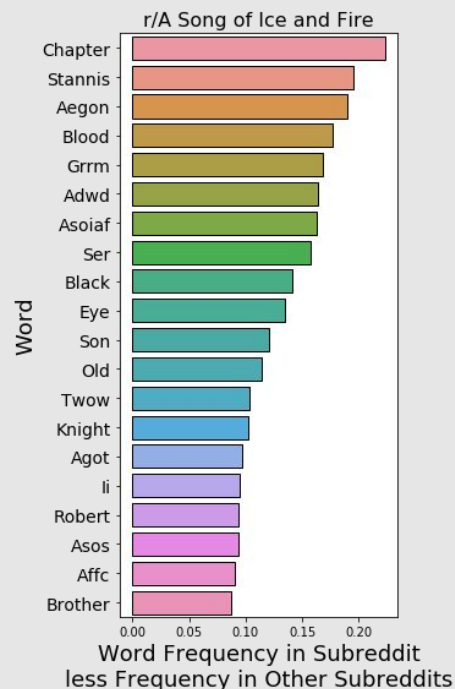
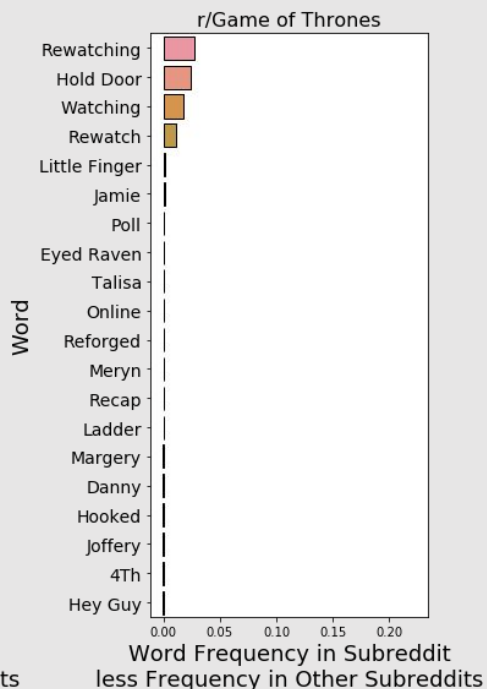Lemma-tize
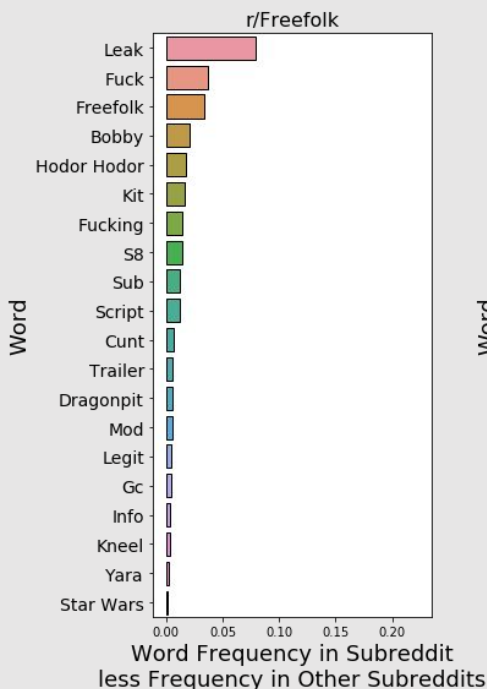
Visualize

Model

Evaluate

Are there Differences?

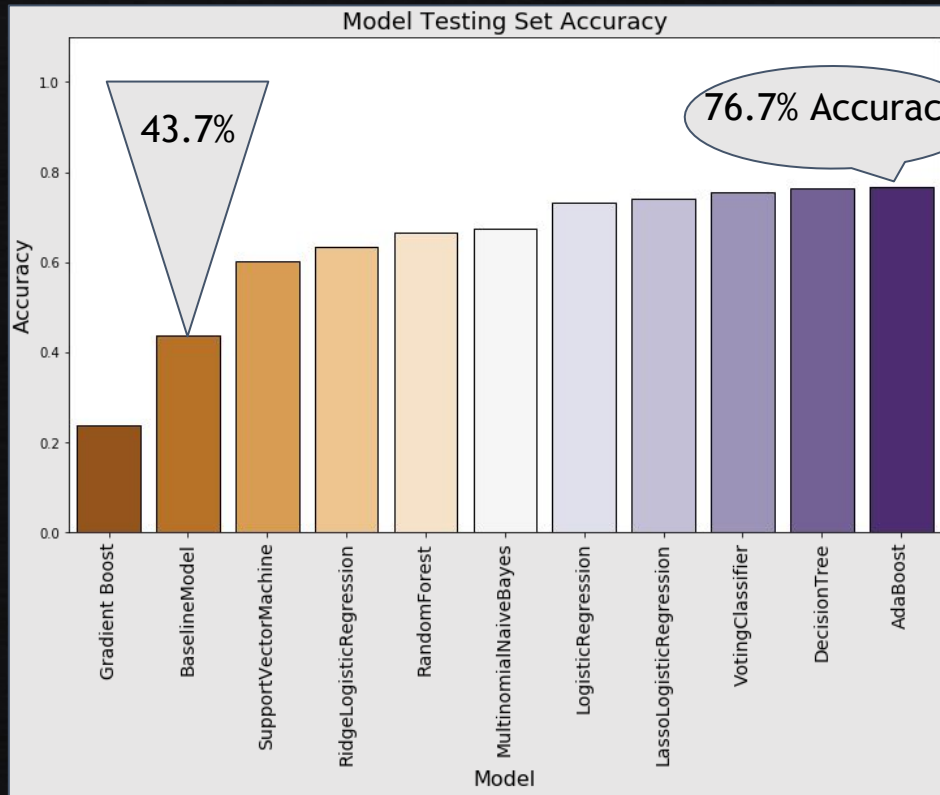Top 20 Most Common Words by Subreddit

Subreddit-Specific Words

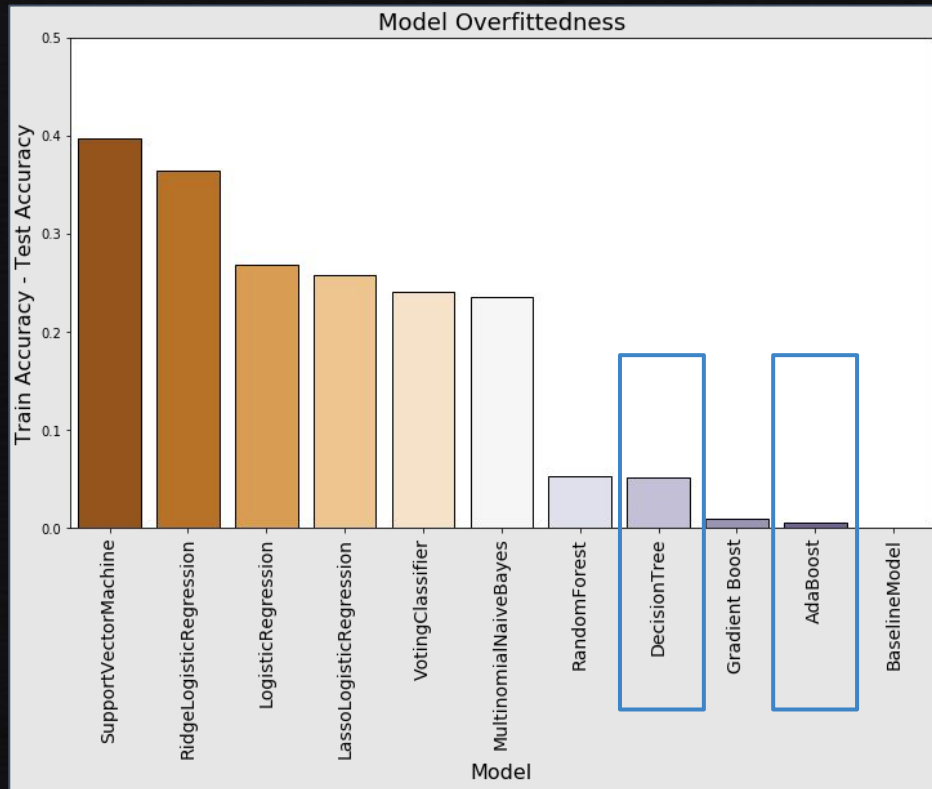

Top 20 Words Most Specific to Each Subreddit

Modeling Results 1

- AdaBoost is our overall best model, with the Decision Tree not far behind.
- Ensemble didn't beat our standalones

Modeling Results 2

Conclusions and Recommendations

- We successfully beat the baseline of 44% accuracy

- r/gameofthrones, r/freefolk, and r/asoiaf have substantial differences and can continue to coexist