

Webscraping

Scrape all songs for 2Pac and John Lennon

Required package: rvest

2Pac: <https://www.azlyrics.com/19/2pac.html>

John Lennon: <https://www.azlyrics.com/j/johnlennon.html>

Text mining for each artist

Required packages: quanteda, stringr

- Character per word
- Most used token, bigrams, trigrams for each singer (calculate the TFIDF value)
- Most used nouns and adjectives
- Lexical diversity
- Readability score (lowest readability song for each artist)

Supervised learning for each artist (Author attribution)

Required package: caret

- Split the data (80% for training the model and 20% for evaluation for each artist)
- NB model for each artist
- SVM model for each artist
- Sentiment trajectory analysis on the two datasets, find the differences between trajectory
- Use q-gram-based similarity calculations and analyse for the two datasets
 - internal similarity (i.e. the similarity among texts of a single singer)
 - Required package: stringdist
 - the similarity across singers (i.e. the similarity between lyrics written by different singers)
- Extract the overall sentiment of each text in your corpus (i.e. a single score for each text).
 - Run a supervised classification model and report detailed performance metrics (i.e. accuracy, precision, recall, F1 score, and area-under-the-curve).