Kate Huangpu
Algorithms
Jonathan Soma
April 6th, 2021

For my classification, I scraped the lyrics of Taylor Swift songs. I scraped the lyrics from around 75 songs, and made each line an entry. I then categorized them by album and by whether the genre was pop or country. I selected words that I knew were popular in her lyrics, both generally as well as those more specific to her pop and country albums.

I tried to use a count vectorizer to select the longer words that are more popular and disregard the very common words. But my count vectorizer ended up only showing one word, "don" which I believe is because of the word "don't." From there I tried to set up my code so that it could predict which lines of songs were pop songs and which ones were country.

Eventually I would like the code to predict from which album the line of lyrica came from, as in my original data set I include that as another column.