Summary:

I created a movie recommendation system using data gathered from The Movie Data Base’s api. Using CountVectorizer, TfidVectorizer, stopwords, CosineSimilarity, and nltk stemming, I was able to use the movie’s description, cast, crew, and keywords to create a fairly accurate recommendation system.

Challenges:

By far the most difficult part of this project was working with TMBD’s api. There was a lot of trial and error with formatting my api requests and sifting through the data to make it usable. A good portion of the data was only accessible through looking up the movie ID, meaning I had to make a bunch of requests for every column I wanted to add, then merge the databases together. Another problem I had was with cleaning the data, as a bunch of data was listed as ID’s and dictionary pairs.

Conclusion:

Although I’m happy with the results, I think the model can be improved upon by taking into consideration popularity, and weighted rating. I also would like to add deep learning to my recommendation system to see if any noticeable improvements are made. For this, I would need a different dataset for labels.