

CECILIA BELL

*MOVIE
RECOMMENDATION
SYSTEM*



WHY?



Improves customer experience



Drives revenue



Resolves cold-start problem



THE DATA

- 34,886 movies
 - 24 origins
-
- Mean plot length: 380 words
 - Max plot length: 6752 words
 - Min plot length: 21 words

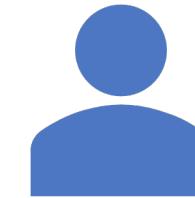
METHODOLOGY



Clean and preprocess text



Build topic model



Create recommendation
system

METHODOLOGY

- **PREPROCESSING:**
 - Tokenizing
 - Lemmatizing
 - Removing proper nouns
 - Vectorizing
- **TOPIC MODELING:**
 - Dimensionality reduction
- **RECOMMENDATION SYSTEM:**
 - Similarity function



FINAL RECOMMENDATION SYSTEM

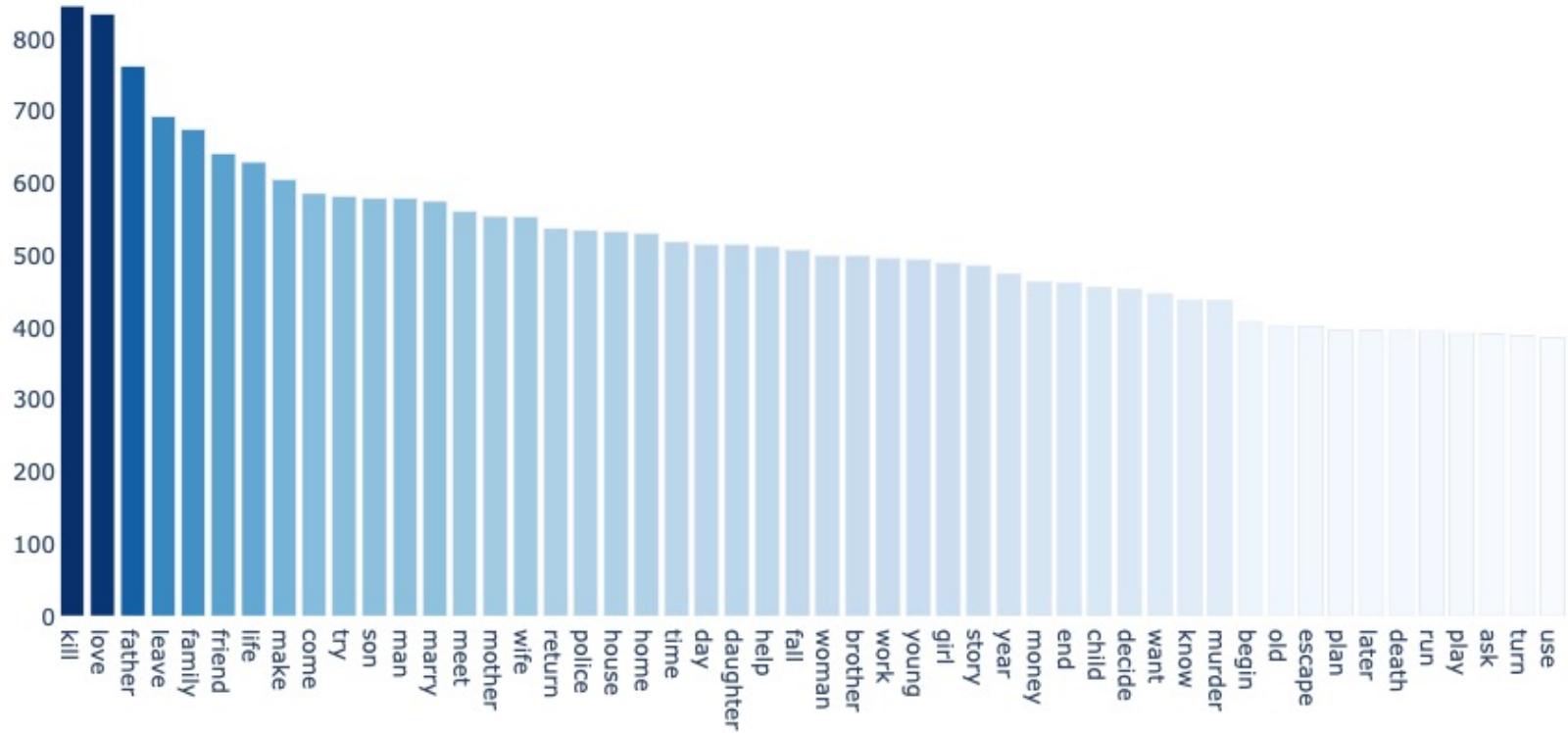
TF-IDF Vectorizer

NMF model with 30 topics

Cosine similarity function

INSIGHTS

Top 50 Most Common Words after Preprocessing



Topic 3: sci-fi

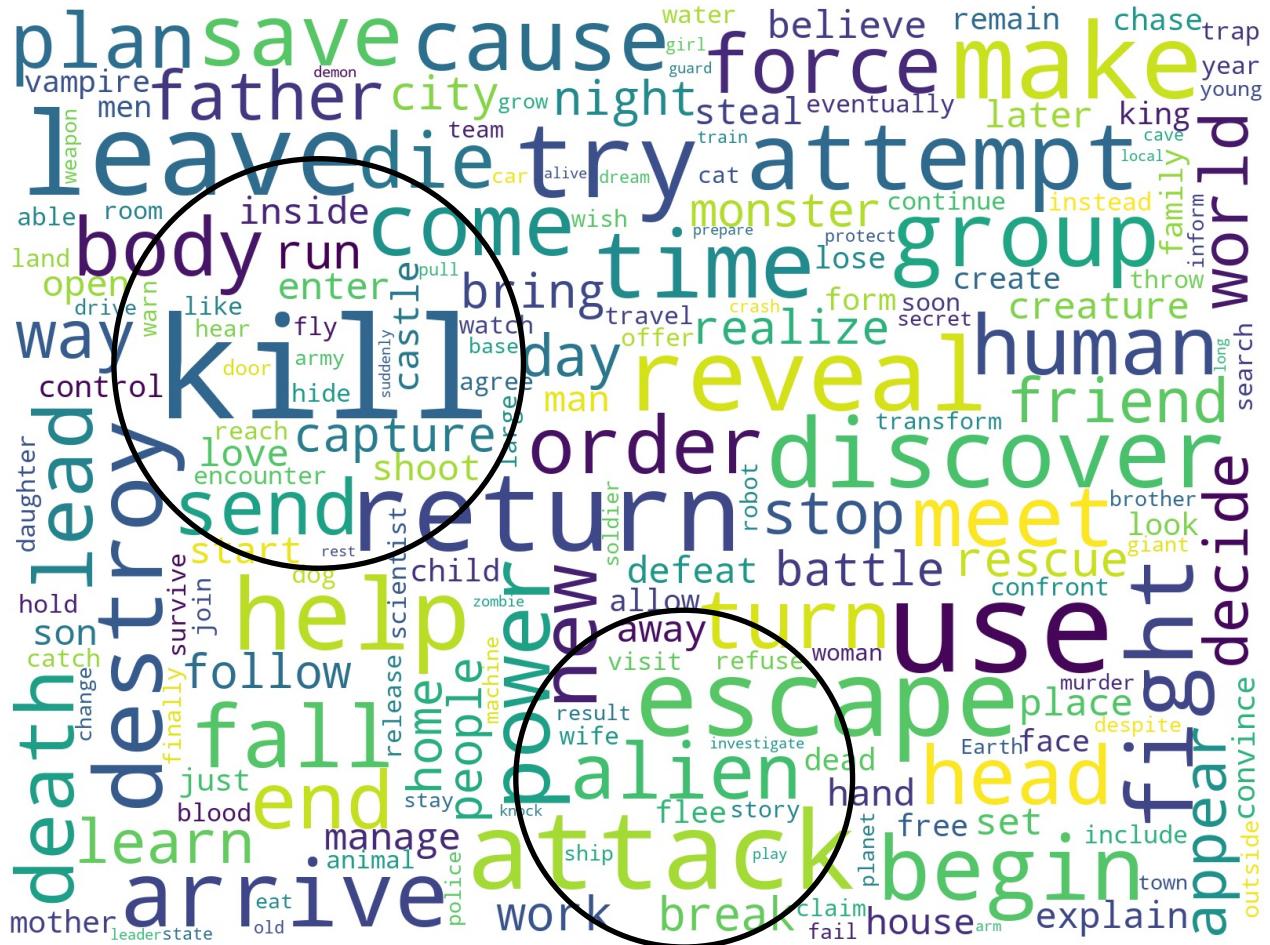
INSIGHTS

Topic 3: sci-fi



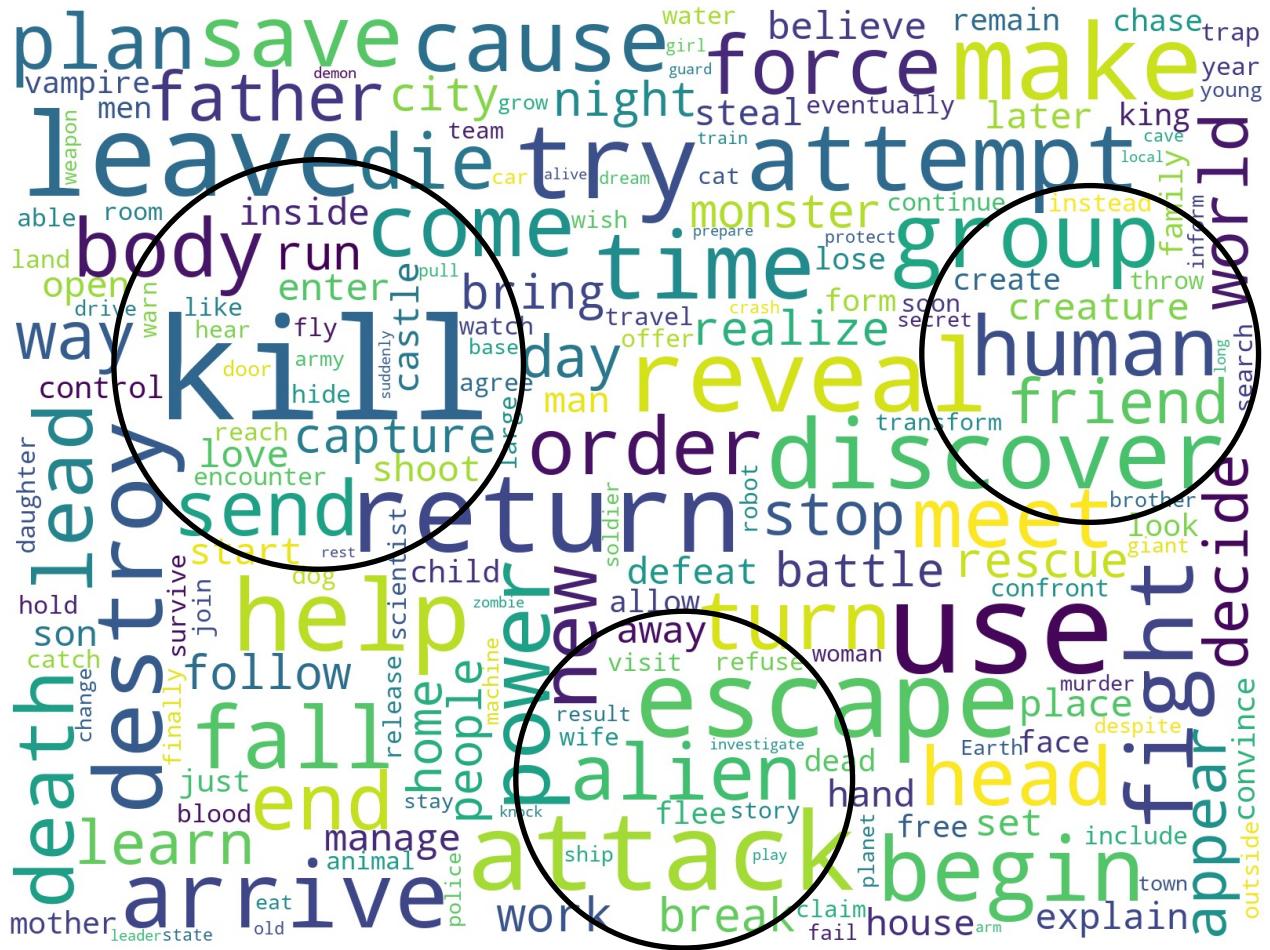
INSIGHTS

Topic 3: sci-fi



INSIGHTS

Topic 3: sci-fi



INSIGHTS

Topic 27: war



INSIGHTS

Topic 27: war



INSIGHTS

Topic 27: war

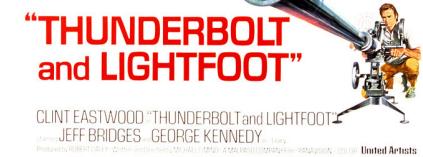
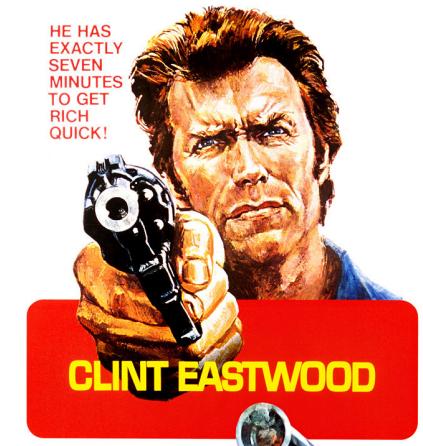
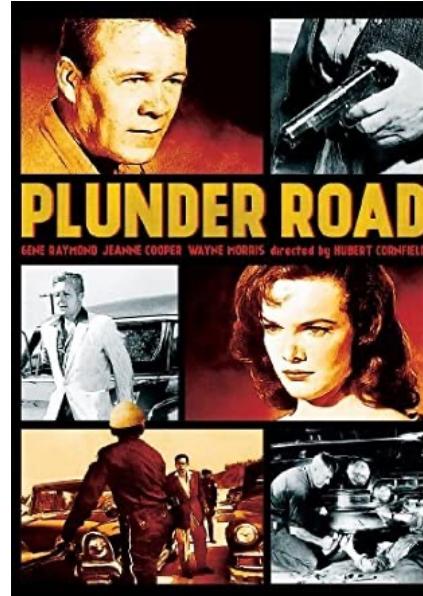


INSIGHTS

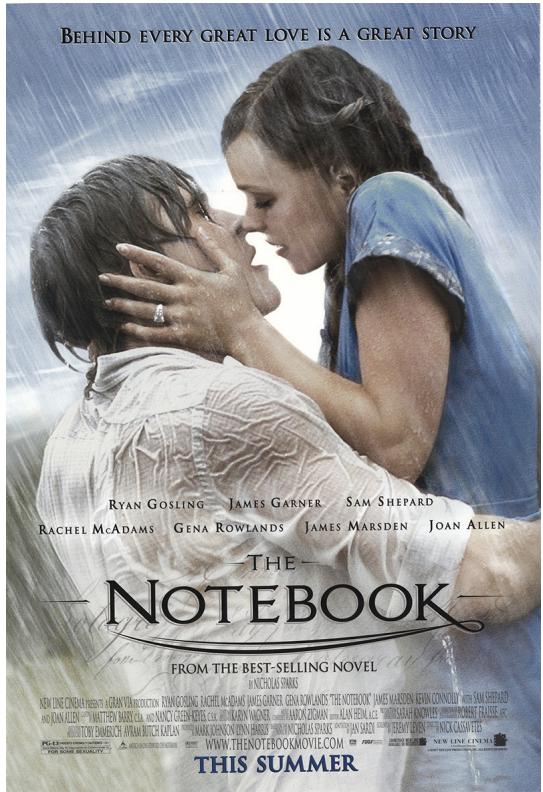
RESULTS



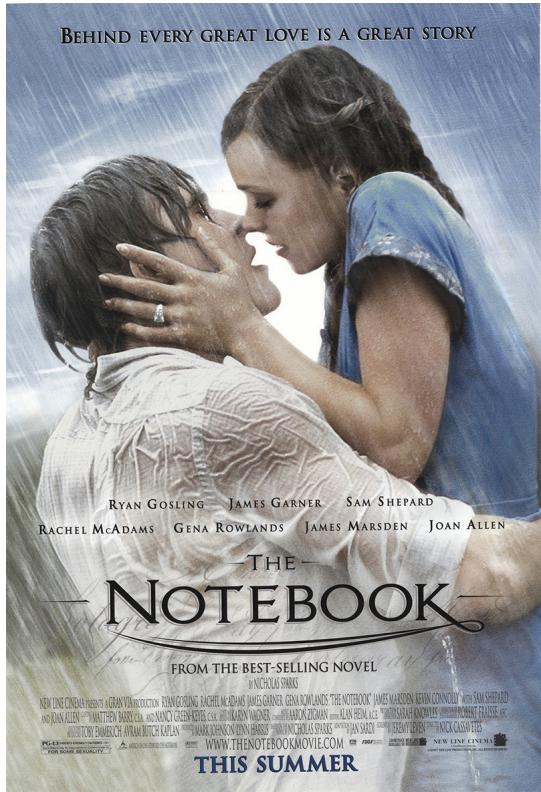
RESULTS



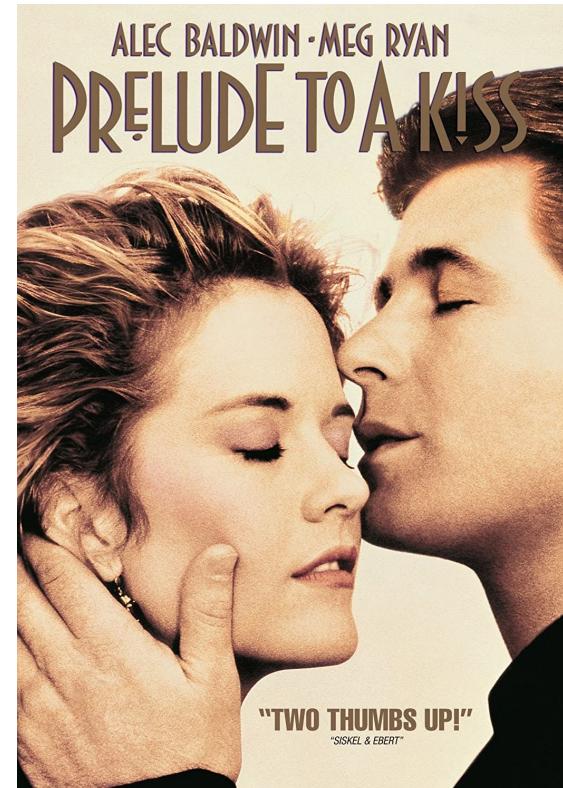
RESULTS



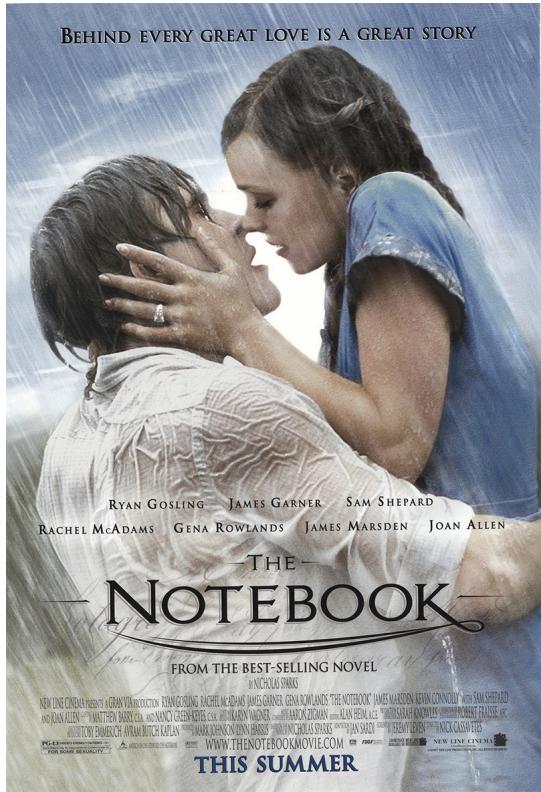
RESULTS



American



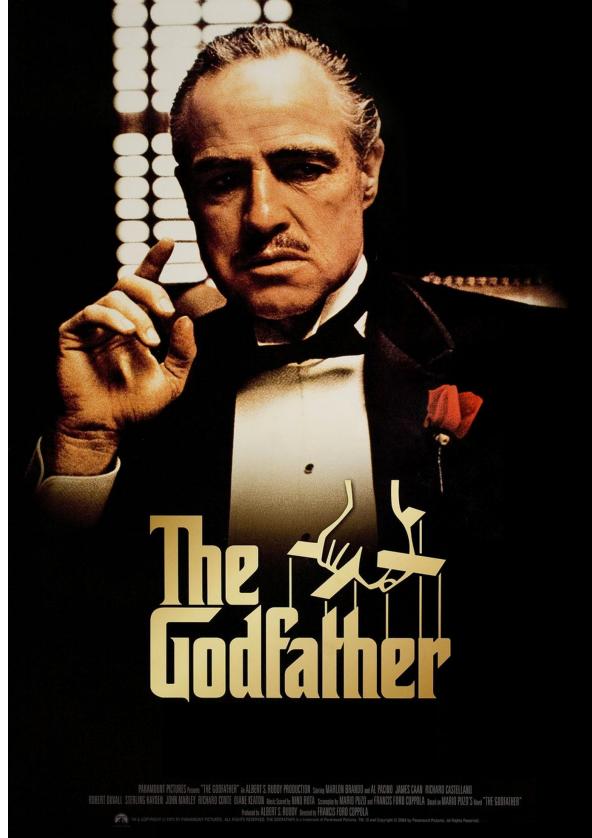
RESULTS



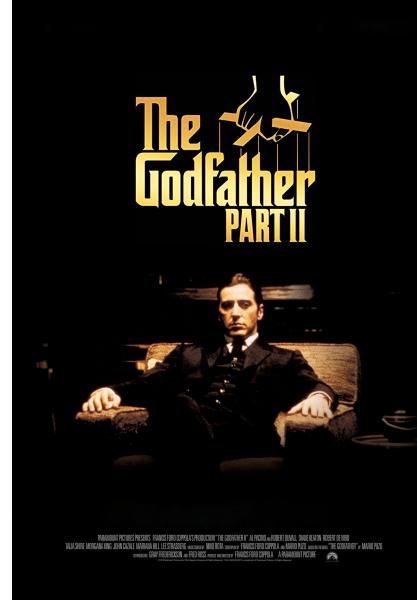
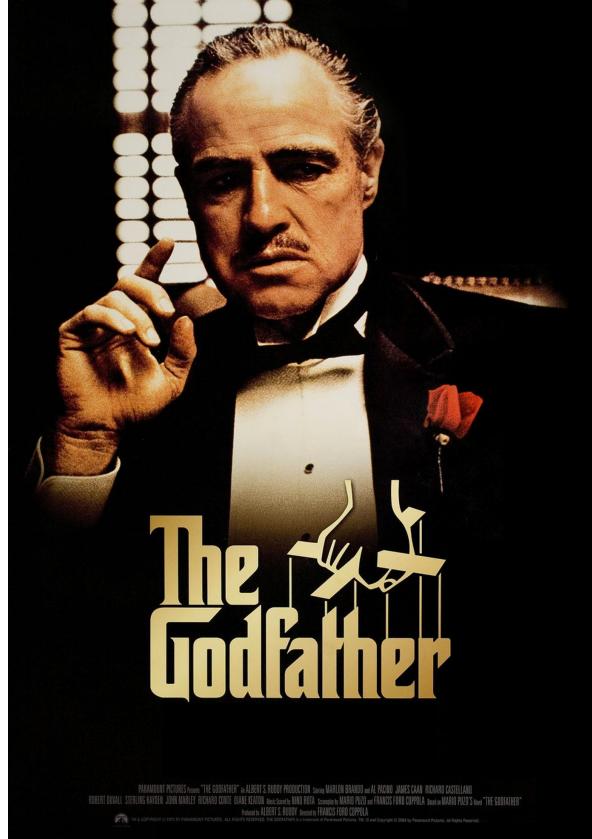
Telugu



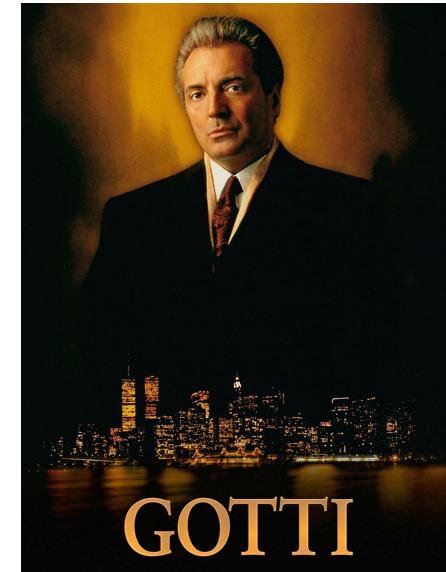
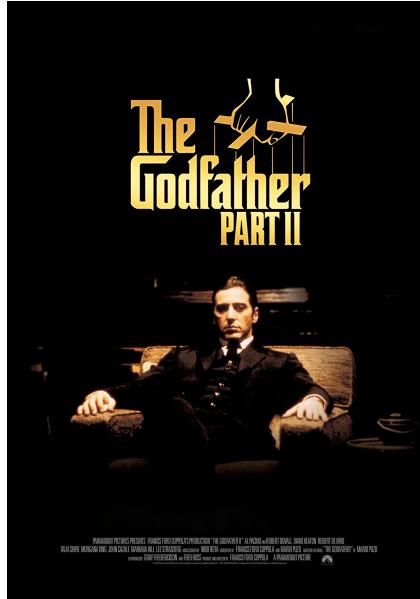
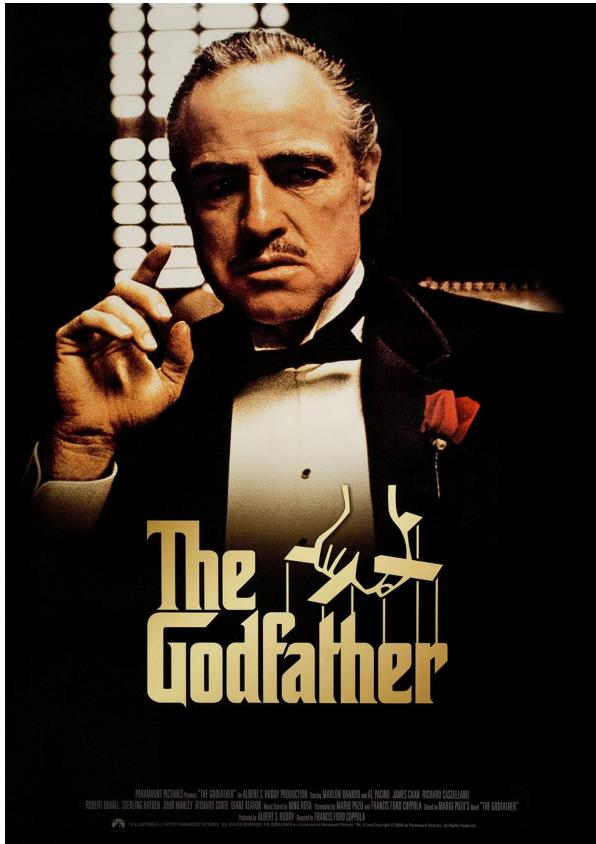
RESULTS



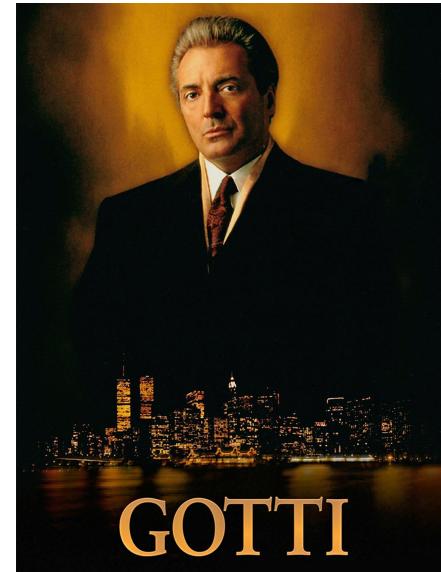
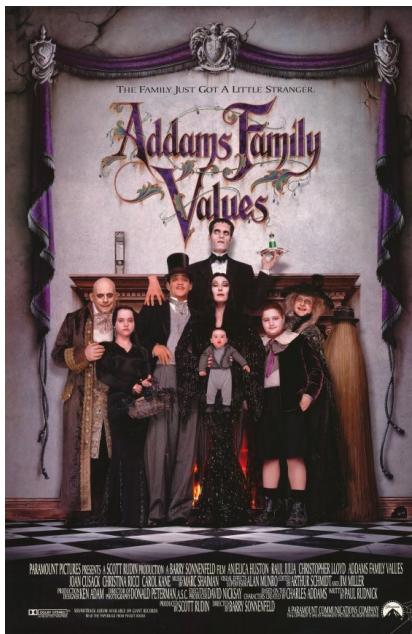
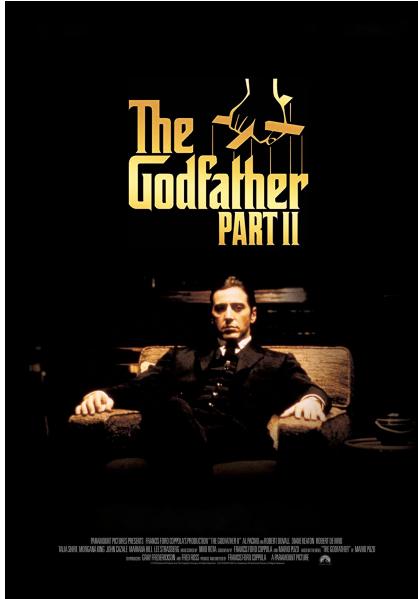
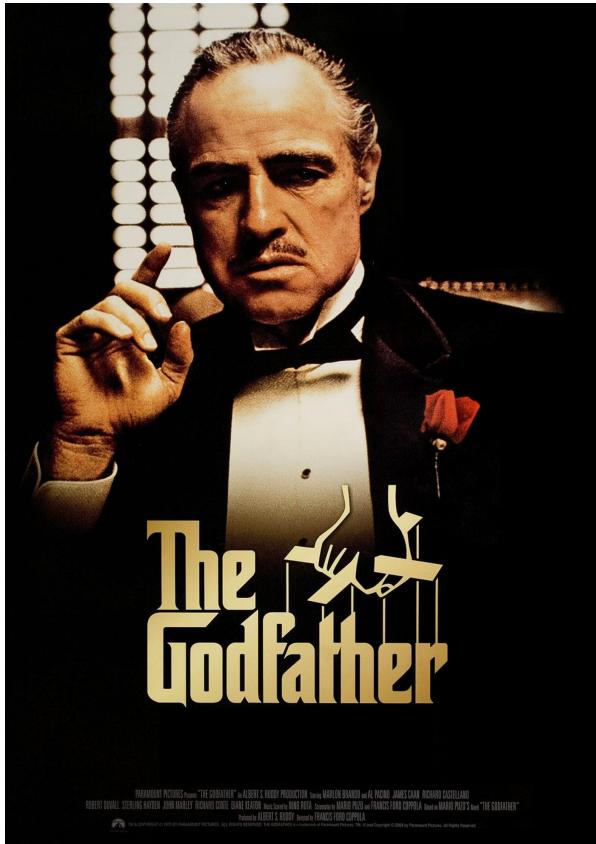
RESULTS



RESULTS



RESULTS

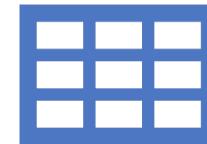


FURTHER WORK



Continue to tune topic
model

Incorporate filters for
year and rating



Include data on director
and cast?