

Movie Mashup: A Content Recommender

By Joshua Stingl

Building a content
recommender
with a Flask
interface.



The Datasets

Kaggle:

Your Machine Learning and Data
Science Community

- IMDb
 - <https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset>
 - Netflix
 - <https://www.kaggle.com/shivamb/netflix-shows>
 - Movie Posters
 - <https://www.kaggle.com/neha1703/movie-genre-from-its-poster>
-

The Process

—

Data combining

1. Determined useful fields
 - a. Description, Title, Genre, Director, Actors, Year, Type
2. Adjust column names to match
3. Add type column to IMDb dataframe
4. Combined into a single dataframe



Text processing

1. Took care of null values
2. Removed spaces from actor and director first and last names
3. Passed descriptions through a translator
4. Lower cased, removed punctuation and stopwords, then lemmatized descriptions and titles
5. Took care of null values
6. Weighted words and combined into a list



Vectorization

1. Passed list of weighted words into a `countvectorizer`
2. Resulting in a 93,642x585,721 sparse matrix with 3,125,341 stored elements



Recommendation

1. Cosine distance

a. Pairwise Distance Chunked



The Flask

