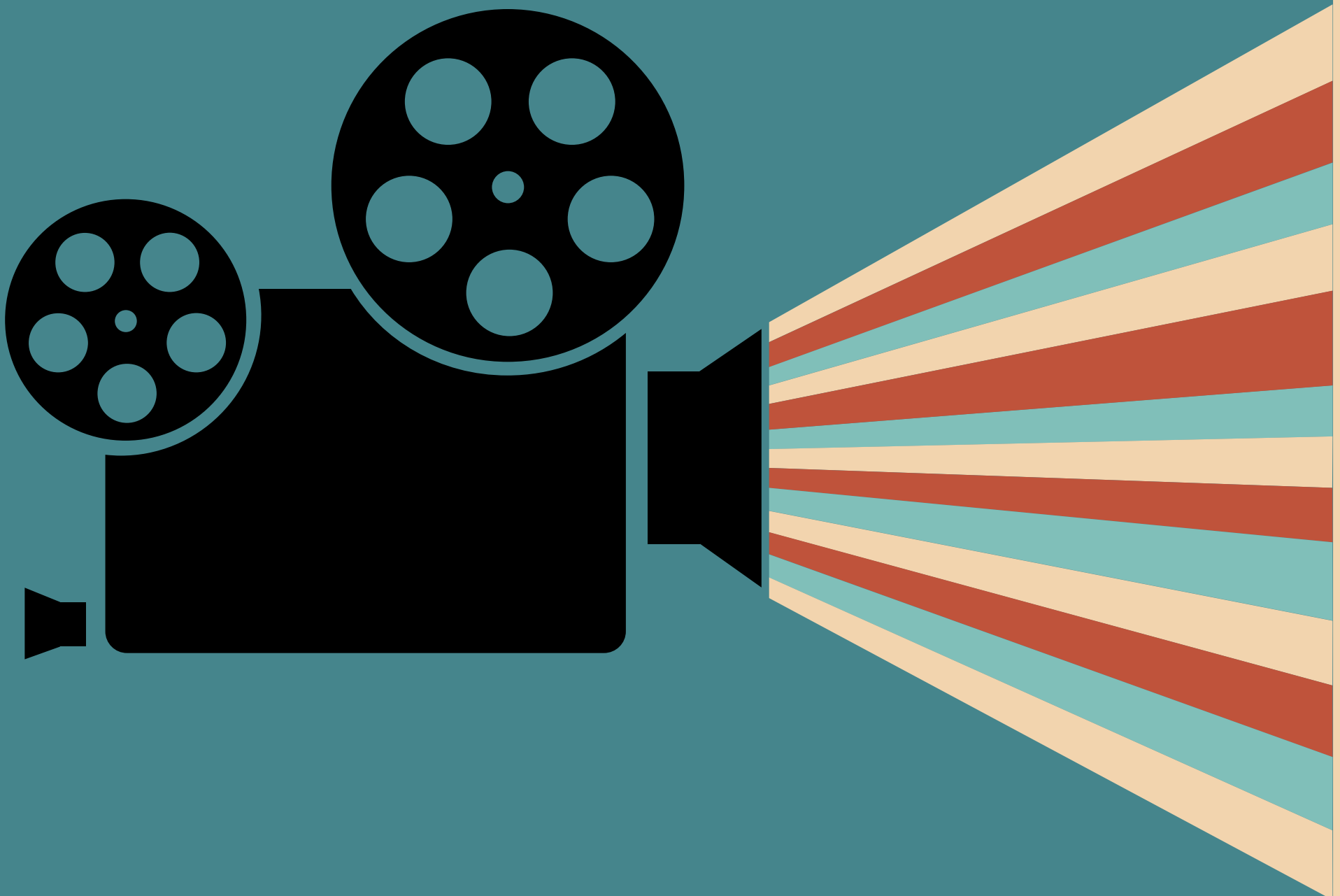


Content-Based Movie Recommendation System

Neshat J. Heravi

JANUARY 26, 2022 • 5:00 PM



The Plot

- Recommendation systems improve user experience and increase revenue
- Content-based recommendation systems can eliminate the “cold-start” problem
- Widely used for recommending movies, articles, restaurants, places to visit, items to buy etc.

Goal

Build a **content-based** movie **recommendation system** based on movie plot overviews using **Natural Language Processing**.

Methodology

Data & Tools

- **45,466** movies featured in the **MovieLens** dataset (movies that are released on or before July 2017) on **Kaggle**

Python Libraries & Packages:

- Pandas & Numpy
- NLTK
- Scikitlearn
- Wordcloud



Text Preprocessing

Cleaning:

- Removing duplicates & missing entries
- Removing extra features
- Removing short overviews

Preprocessing:

- Removing proper nouns and numbers etc.
- Tokenizing
- Parts of speech labeling
- Lemmatizing
- Vectorizing (TF-IDF)



Topic Modeling & Recommendation System

Topic Modeling:

- Dimensionality Reduction
- Non-negative Matrix Factorization (NMF)

Recommender System:

- Topic Profiles
- Cosine similarity

Insights

Wordcloud Text Visualization for Different Topics:



Results

Content-based Recommendation System:

- **Input:** Movie Title
- **Output:** 5 Movie Recommendations

```
recommendations('Bonnie and Clyde')
```

```
['The Foreigner',  
'The Interrupters',  
'Small Town Gay Bar',  
'Ivanhoe',  
'The Rosa Parks Story']
```

```
recommendations('Robin Hood')
```

```
['Mulberry Street', 'San Pietro', 'The Speak', 'Rustlers' Rhapsody', 'Kill!']
```

Example

Input: Bonnie & Clyde

Output: ['The Foreigner', 'The Interrupters', 'Small Town Gay Bar', 'Ivanhoe', 'The Rosa Park Story']

Content Theme: Resistance , Rule-breaking & Interruption

Conclusion & Future Works

- The created recommendation system is based on movie overview plot contents.
- It takes in a movie title and will suggest 5 other movies

As Is: 'Zorro' and 'Shrek' have similar plots!
They really do!!

```
recommendations('Zorro')
```

```
['Spies', 'Guns', 'House of Flying Daggers', 'Shrek', 'Casablanca Express']
```

How to Improve the System?

By adding filters such as movie origins, genre, release date etc.

Thank You!

