

CSE 5334 – 003 Data Mining

Term Project Proposal

Project: Hall of Movies

Project Idea: I will be developing a web application which can search and classify movies based on what user input. The search which will be done by taking some textual inputs from the user and classification will be done on various features present in the dataset. Aim is to have a web application that can do following things:

- Search text information
- Build a classifier
- Build an image captioner

Dataset: The dataset is from Kaggle: <https://www.kaggle.com/beyjin/movies-1990-to-2017#Movie> Movies.csv

The dataset has Movie's name, language, runtime, and description of the movie plot, its IMDB ratings and many more attributes.

Features: It will provide following features:

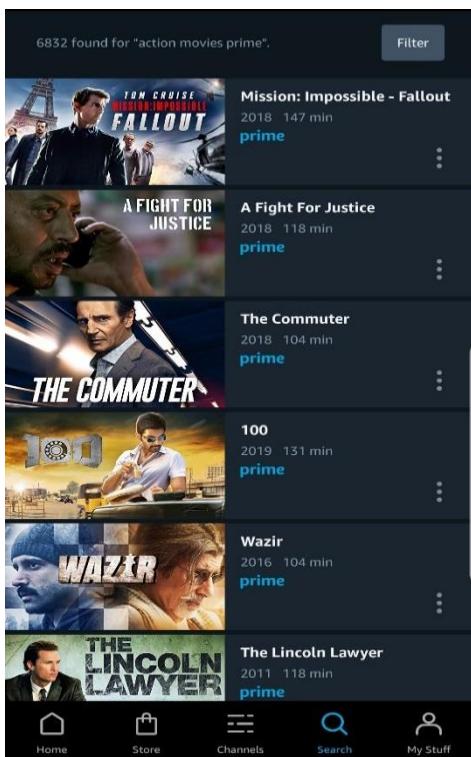
- The application will classify motion movies based on ratings, similar interests etc.
- Search for the movies with free text.

Why?

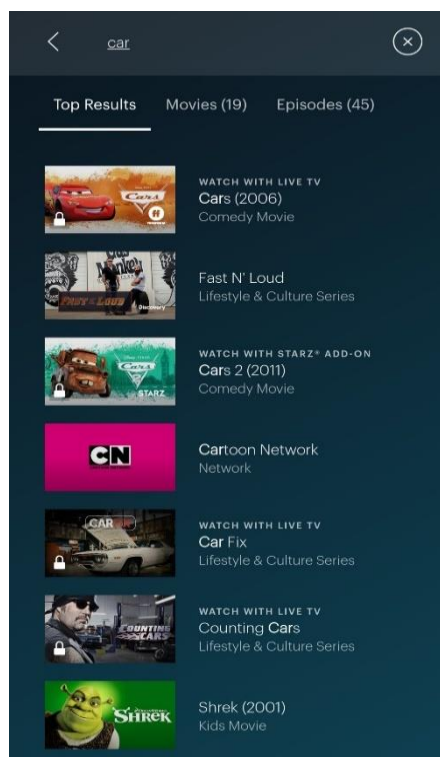
The idea is to make an application which can help you to find movies when you search with some free text related to that movie. This will help user to find the movies they like even if they don't remember the movie. Another important functionality is to show movies with similar interests of the user based on what movie they watched. This will be depending on the ratings, genre, type of the movie they recently viewed. This can help them to watch another movie which will most likely be of their interest.

Some Real-World Applications:

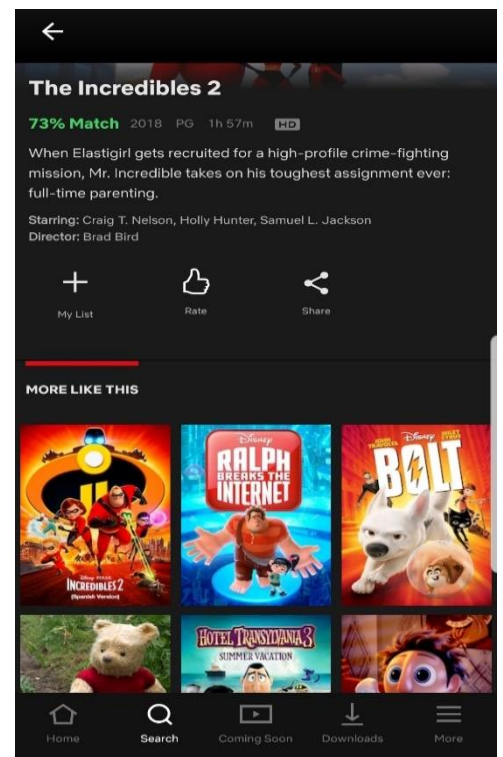
1. Netflix
2. Amazon Prime
3. HBO
4. Now Hulu



SEARCH BASED ON GENRE



SEARCH BASED ON FREE TEXT

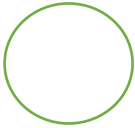


MOVIES BASED ON WHAT YOU WATCHED

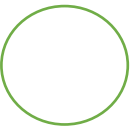
TEXT SEARCH

Car Racing

Search



Fast and Furious - 1



Rush

A Murder Scene

Classify

Action: 60%

Thriller: 20%

Mystery: 20%