

CSE 5334 – 003

Data Mining Term Project Proposal (Fall 2019)

Yash Bardapurkar

1001731650

Movie Search and Classifier

Introduction

The project for the course is a web application to search for movies by keywords matching from the synopsis, classifying a movie by genre based on the description given by the user, and image captioning.

Dataset

The dataset contains metadata for 45,000 movies listed in the Full MovieLens Dataset. The dataset consists of movies released on or before July 2017. Data points include cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries etc.

<https://www.kaggle.com/rounakbanik/the-movies-dataset>

Features

- Movie Search: Search for a movie by using simple queries given by the user
- Movie Classification: Classify movies based on the description given by the user
- Image Captioning: Generate a caption for the image uploaded by the user

Why this app

We often spend a lot of time trying to find a movie when we do not know the title of the movie, but we remember just small key points from the plot of the movie. This application makes it easy for users to search movies by taking simple text queries from the users, and returning the best results. This application also lets the user classify the movie by genre based on the description provided by the user.

Similar applications

Search engines of:

1. Netflix
2. Hulu
3. Amazon Prime Video

Sketches

1.

Movie Search

Enter Keywords to search

1. The Fast and the Furious	
(2001)	IMDb 6.8
2. Baby Driver	
(2017)	IMDb 7.6

2.

Movie Classification

Enter Keywords

Action	80%
Thriller	60%
Drama	20%

3.

Image Captioning

Upload Image

Drag and
drop image
here

Caption: