**Movie Recommendation System:**

based on :-

1. Movie title

2. Movie overview

3. Movie genre.

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

**Approach :**

Using Transfer Learning Based on Google’s Deep Learning Model BERT

(Bidirectional Encoder Representations from Transformers) to convert the overview of the movie into Sentence embeddings .

Pretrained BERT Model : wwm\_uncased\_L-24\_H-1024\_A-16 is used to convert each token of the sentence to 1024 dimension embedding ,

features are extracted from the last 4 hidden layers and mean of them is taken to get a token embedding of shape = (seq\_length,1024) (seq\_length=256). Now average of all the token embeddings are taken of the sentence to get the sentence embeddings.

Now , Cosine Similarity is used to give high score to similar sentence embeddings (movie’s overview).

Top ten movies with highest similarity score is recommended.

**Features :**

1. Enter movie title , it ‘ll find the movie overview and return the similar movies by overview.
2. Enter the description of the movie it’ll extract the sentence embeddings and get the similar movies .

**Future work:**

1. Comparing overviews of movies with similar genres only , to give better results in less time as same overview of movies can give different sense in different genres ( comedy , action etc.)
2. Selecting top rated 10k movies to recommend .
3. Topic modelling

Mistakes:

1. Movies with same name like (“The Mask”,”The Avengers” etc) are overwritten while extracting features from the model. Hence while embeddings of these movies are same and are not giving expected result . Solution change the Name and extract embeddings.

**Didn’t work out:**

1. The Overview data of the movie is unstructured and hence Word vectorizer/Tfidf method didn’t work good.

**Other Approaches:**

1. Training a Seq2Seq model or Sequence model on Overview Data to predict multi label genres and using it to extract features and find similar movies by cosine similarity.
2. Fine-Tuning a Bert model to predict genres and extracting features from it.

**Project Setup :**

1. Unzip the code files.
2. Create a folder “tmp”.
3. Install bert pretrained model from :

<https://storage.googleapis.com/bert_models/2019_05_30/wwm_uncased_L-24_H-1024_A-16.zip>

Unzip the model in a sub-directory “tmp/”.

1. Download the embeddings from : <https://drive.google.com/open?id=1TvYsNERvBwjGrxPE9ytTtgml776nXJQr>
2. Download the data and unzip in the main folder :

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

Directory\_overview:

-movie\_recommendation(main folder)

-/emb.pickle

-/tmp (model folder)

-/Bert pretrained model unzipped

-/movie dataset (unzipped)

-/code files (unzipped)

To see the result run recommender.ipynb

To start the api run server.py