**Movie Analysis**

**Functionality of code or Components of Code**

All the components which were mentioned in the “Idea” file are implemented without any changes.

1. **Data Pre-Processing**
   1. The dataset given, contains a lot of un-processed information. Using Python library **“Pandas”** we have converted the datasets into inferences.
   2. The inferences are then stored into another dataset including all the previous data.
   3. It is to be noted that the entire data used is solely from the dataset given.
2. **Image Input format**
   1. The query from the user is in the form of image. The image is then converted to four different formats like RGB and grayscale etc. Using this combination we detected the text in the image differentiating it with the background using edge-detection. **Open-cv** is used for the above.
   2. Now, using **Pyteserract** from the converted image, text is extracted and then an array of all required constraints and keyword parameters are created and sent for processing/visualization.
3. **Text Input format**
   1. The query from the user is in the form of text. The text is converted into vector assigning different values using **Tensorflow** (Tokenizer) simulating Natural Language Processing(**NLP**).
   2. The converted vectors are then matched with pre-defined keywords and constraints to extract the values as parameters for sending it to processing/visualization
4. **Data Visualization**
   1. The converted input which is generated using Image and Text processing is used for fetching the data from the datasets.
   2. In some cases, Output should be just a simple sentence. These cases have been taken into consideration.
      1. Ex
         1. Query: Who is the lead role in the movie 3 idiots.
         2. Answer: The lead actor is “Rancho”
   3. In some other cases, Output should be in the form of graphs. Consider the example “What is the variation of values of emotions over the years since 2007?”
   4. Complex queries are also handled, where multiple queries can also be asked by the user.
      1. **Ex.** What is the genre of the movie Shamitabh and who is the lead role in the movie 3 idiots and what is the plot of the movie black mail.
5. **Prediction using Machine Learning**
   1. Firstly, we created a graph from the emotions dataset which is used to show the variations in emotions over the years from 2007 to 2017
   2. Given the values of different categories in emotion (Angry, Sad, Happy, Disgust, Surprise, Neutral, Fear) by the user we can predict the year in which the movie released or if it is relatable to the audience in the present scenario.
   3. We have used “**SciKit-Learn**” to predict the value of year.