**Name of the Problem Statement** : Movie dataset analysis

**Team Size and members** : 4 Members

* Nikhil J.S.K
* Prasanth Gelli
* Anil Raj T.
* Raja Rahul M.

**Problem Statement**

Upon exploring the dataset provided, we have found that the data which is given is quite unorganized and impossible to understand the results and inferences in a whole. So, we’ve made a command line based interface using Python-3 which is capable of asking the query from user in text format or even in Image format and get the answer from the dataset. We’ve used Natural Language Processing and Image processing libraries to achieve this. We have also analyzed the emotions of the film in each year and then could predict the year given the rate of emotion. The solution contains different types of graphs and text outputs to give a holistic comprehensibility to the end user.

**Scope of Work**

* Text Processing
  + Since the first component in the solution is to analyze the text given by the user, we have used tensorflow libraries to convert the text in form of vectors, find the relevant keywords and constraints to the solution and then create an array with the possible outcomes.
* Image Processing
  + To analyze the image and extract text from it, we have used different libraries like “Pytesseract” and “Open-cv”. We have then converted them to vectors and just treated them as text format.
* Data Pre-Processing
  + The dataset given has a lot of incomprehensible inferences, so we have used Python library “Pandas” to reframe all the needed data and then create a new dataset from the given dataset.
  + While processing this data we have extracted some information from the existing and then used them. We have NOT used anything out of the given dataset.
* Data Visualization
  + The most important part of the entire solution is to make the user understand the result of his/her query. We have used libraries like “Matplotlib” and “Seaborn” to achieve some spectacular results which are comprehensible too.
  + Data visualization could also be in the form of simple text for questions like “Who is the lead role in the movie Shamitab?”. Queries of such type have also be take care of.
* Prediction using Machine Learning
  + Firstly, we have processed the information given in emotion dataset, created a graph which displays the variation in each category of emotion year-wise.
  + Then we have used Linear Regression algorithm to predict the possible year for given values of emotion in each category(Angry, Sad, Happy, Disgust, etc.). The library used here is “Scikit-Learn” from Python.

**Technologies**

The project was made using Python 3.

* Input Libraries
  + Tensorflow
  + Pytessearact
  + Open-cv
* Data Pre-Processing Libraries
  + Pandas
  + Numpy
* Data Visualization Libraries
  + Matplotlib
  + Seaborn
* Machine Learning Libraries
  + SciKit-Learn

**Role of Team Members**

1. **Nikhil J.S.K -** Data Processing and Visualization
2. **Prasanth G. -** Text and Image Processing
3. **Anil Raj T. -** Text and Image Processing
4. **Raja Rahul M. -** Data Processing and Visualization

Thank you