NAME – PRITHA MANDAL

AFAME TECHNOLOGIES

DATA SCIENCE INTERNSHIP

MOVIE RATING PREDICTION WITH PYTHON

Objective -

1.Building a model that predicts the rating of a movie based on features like genre, director, and actors. Have can use regression techniques to tackle this problem.

2.The goal is to analyse historical movie data and develop a model that accurately estimates the rating given to a movie by users or critics.

3.Movie Rating Prediction project enables you to explore data analysis, preprocessing, feature engineering, and machine learning modelling techniques. It provides insights into the factors that influence movie ratings and allows to build a model that can estimate the ratings of movies accurately.

**Indexs**

'Name', 'Year', 'Duration', 'Genre', 'Rating', 'Votes', 'Director','Actor 1', 'Actor 2', 'Actor 3'

### Step 1

### Importing of required libraries

1. Pandas
2. Numpy
3. Matplotlib
4. Seaborn
5. Plotly

\*\* (At the time of feature engineering 1 more library is needed that is scikit-learn)

### Step 2

Mounting the drive to get the data set load and then read the dataset.

### Step 3

##### Data Cleaning

Checking for null value is present in the dataset or not. Then get the information about teh dataset. Then checking for duplicate value. Then dropped the null values and the duplicate values. Then get the names of the columns.

### Step 4

##### Data Pre-processing

Replacing the unwanted brackets from 'Year' index entities. Then 'Duration' column values are changed from string data type to numeric data type. Convert unique values of the Genre column and set mode.

### Step 5

##### Data Visualization

1. To get a visualized information about the data, we are taking the **Histogram** where in x axis Year is showing and in y axis Probability Density is showing.
2. Line graph has been used to get the visualize information about the relation between Average Rating by Year And Genre.
3. Using Histogram, The Distribution of Rating has been shown where in x axis Rating and in y axis Probability Density has been shown.

### Step 6

##### Feature Engineering

1. Importing essential modules from scikit-learn.
2. Dropping the Name of the movies because it is not affecting the prediction.
3. Grouping the columns with their average rating and then creating a new features.
4. Keeping the predictors and target variable.
5. Spliting the dataset into traing and testing part.

### Step 7

##### Model Building

1. Building the linear regression model and train it.
2. Getting the evaluation through logistic regression.

### Step 8

##### Model Testing

1. For testing, we create a new dataframe with values close to the any of our existing data to evaluate.
2. Predict the movie rating by entered data.
3. Display the predicted result.

So we are getting a satisfying prediction.