**Movie Rating Prediction**

**Objective:-**

1. Explain the purpose of the project: Predict movie ratings based on user and movie data.
2. Briefly outline the methods used: Machine learning algorithms like regression, collaborative filtering, or neural networks.

**About DataSet :**

1. **Name: Movie title.**
2. **Year: Release year (some entries are missing).**
3. **Duration: Duration of the movie in minutes (many entries are missing).**
4. **Genre: Movie genre(s) (e.g., Drama, Comedy).**
5. **Rating: IMDb rating (target variable, many entries are missing).**
6. **Votes: Number of votes the movie received.**
7. **Director: Movie's director.**
8. **Actor 1, Actor 2, Actor 3: Leading actors in the movie.**

**1.Data Cleaning and Preprocessing:**

* **Handled missing values by dropping rows with null values in key columns like Rating, Director, and actor columns.**
* **Transformed the Votes, Year, and Duration columns into numeric formats suitable for analysis.**
* **Filled missing values in the Duration column with the median.**

**2.EDA (Exploratory Data Analysis):**

* **Analyzed the top 10 movies by rating.**
* **Visualized:**
  + **Distribution of top movies with respect to year and genre.**
  + **Top directors based on average ratings.**
  + **Relationship between votes and ratings.**
  + **Top actors by the number of movies.**
  + **Yearly trends in the number of movies released.**
  + **Duration trends over the years.**
* **Highlighted movies with ratings greater than 8 and votes exceeding 10,000.**

**3. Feature Engineering:**

* **Create additional features from existing data, such as:**
  + **The number of genres per movie.**
  + **Actor or director popularity based on their average ratings or movie count.**
  + **Binary indicators for specific popular genres (e.g., is\_comedy, is\_drama).**
* **Normalize numeric features like Votes and Duration if they have large value ranges.**

**4. Model Building:**

* **Prepare the dataset for machine learning:**
  + **Separate features (Year, Duration, Genre, etc.) and target (Rating).**
  + **Use one-hot encoding or other techniques for categorical variables like Genre and Director.**
* **Experiment with different regression models:**
  + **Linear Regression**
  + **Random Forest Regressor**
* **Split the dataset into training and testing sets for evaluation.**

**5. Evaluation Metrics:**

* **Evaluate model performance using metrics like:**
  + **Mean Squared Error (MSE)**
  + **Mean Absolute Error (MAE)**
  + **R² Score**
* **Use cross-validation to ensure robust results.**

**6. Visualization of Predictions:**

* **Compare actual vs. predicted ratings using scatter plots.**
* **Plot feature importance to identify key contributors to movie ratings.**

**7. Insights and Reporting:**

* **Summarize key findings and insights from EDA and modeling.**
* **Source Code –**

**https://github.com/Adarsh9154/Movie-Rating-Prediction**