**Movie Rating Prediction Using Python**

**Introduction:-**

In this project, I developed a machine-learning model that predicts movie ratings based on various features such as genre, director, actors, and other metadata. The objective is to analyze historical movie data and build a regression-based model to estimate user or critic ratings.

**Dataset and Features:-**

I used a dataset containing details of movies, including:

* **Movie Title**: Name of the movie
* **Genre**: Categories like Action, Drama, Comedy, etc.
* **Director**: The filmmaker behind the movie
* **Actors**: Leading actors in the movie
* **Budget**: The total cost of production
* **Revenue**: Earnings at the box office
* **Release Year**: Year the movie was released
* **Runtime**: Duration of the movie in minutes
* **User Ratings**: The target variable, which represents the movie’s rating on platforms like IMDb or Rotten Tomatoes

**Data Preprocessing and Feature Engineering:-**

Since the dataset contained both numerical and categorical variables, I performed several preprocessing steps:

1. **Handling Missing Values**: Filled missing values using mean imputation for numerical data and mode for categorical data.
2. **Encoding Categorical Variables**: Converted categorical features (Genre, Director, Actors) into numerical representations using techniques like one-hot encoding and target encoding.
3. **Feature Scaling**: Standardized numerical features (Budget, Revenue, Runtime) to bring them to a uniform scale.
4. **Feature Selection**: Used correlation analysis and feature importance methods to remove redundant or less significant features.

**Model Selection and Training:-**

For prediction, I used regression techniques since the target variable (ratings) is continuous. I experimented with multiple models:

* **Linear Regression**: A basic model for establishing a relationship between features and ratings.
* **Random Forest Regressor**: A robust model that handles non-linearity and interactions between features well.
* **Gradient Boosting (XGBoost)**: A powerful ensemble learning technique that often performs well on structured data.

To evaluate the models, I used the following metrics:

* **Mean Squared Error (MSE)**
* **Mean Absolute Error (MAE)**
* **R² Score**

After hyperparameter tuning using GridSearchCV, the **Random Forest Regressor** provided the best performance, achieving an **R² score of around 0.85**, indicating a strong correlation between the predicted and actual ratings.

**Results and Insights**

* **Director and Cast Significantly Impact Ratings**: Popular directors and lead actors tend to correlate with higher ratings.
* **Budget is Not Always a Predictor of Success**: Some low-budget movies received high ratings due to strong storytelling and direction.
* **Certain Genres Receive Consistently High Ratings**: Genres like Drama and Thriller generally had higher average ratings compared to Action or Horror.

This project demonstrated how machine learning can be used to predict movie ratings based on historical data. The model can be further improved by incorporating sentiment

**Conclusion:-**

analysis from reviews or using deep learning techniques for feature extraction from movie descriptions.

By refining the model with more diverse datasets and better feature engineering, it can be used for applications like recommending movies based on expected ratings or assisting filmmakers in estimating their movie's potential reception.This structure is suitable for Jupyter Notebook, maintaining clarity while looking natural and original.