**Movie Rating Prediction using python**

**Project: Movie Rating Prediction**

**Overview**

I developed a machine learning model to predict movie ratings based on features such as genre, director, and actors using historical movie data. This project involved various stages, including data cleaning, feature engineering, exploratory data analysis, model selection, evaluation, and prediction using Python and key data science libraries. The goal was to build an accurate and efficient predictive model to provide insights into the factors influencing movie ratings.

**Problem Statement**

The objective of this project was to predict the ratings of movies based on their attributes, such as genre, director, and cast, to better understand the elements that contribute to higher ratings.

**Tools & Technologies**

Python: The primary programming language used.

Pandas & NumPy: For data manipulation and analysis.

Scikit-Learn: For building and evaluating machine learning models.

Matplotlib & Seaborn: For data visualization and exploratory data analysis.

**Data Preparation**

Data Cleaning: Handled missing values and inconsistencies in the dataset. Dropped rows with missing 'Rating' values and filled other missing values with 'Unknown'. Cleaned the 'Votes' column to remove commas and converted it to float.

Feature Engineering: Created meaningful features from the existing data:

Duration: Removed 'min' from the 'Duration' column and converted it to float.

Numerical and Categorical Features: Standardized numerical features and one-hot encoded categorical features (Genre, Director, Actor 1, Actor 2, Actor 3).

**Exploratory Data Analysis (EDA)**

Visualized distributions and relationships between features using histograms, box plots, and heatmaps.

Identified key factors affecting movie ratings, such as genre, director, and actors.

**Model Development**

Algorithm Selection: Evaluated multiple machine learning algorithms including:

- Random Forest Regressor

Preprocessing Pipeline: Used a `ColumnTransformer` to preprocess numeric and categorical features. Applied standard scaling to numerical features and one-hot encoding to categorical features.

Training: Split the data into training and testing sets and trained the model using a `RandomForestRegressor`.

Model Tuning: Used cross-validation and grid search to optimize hyperparameters and improve model performance.

**Evaluation Metrics**

Mean Absolute Error (MAE): Measures the average magnitude of errors in a set of predictions.

Mean Squared Error (MSE): Measures the average of the squares of the errors.

Root Mean Squared Error (RMSE): Square root of the MSE.

R-squared (R²) Score: Represents the proportion of the variance for the target variable that's explained by the independent variables in the model.

Mean Absolute Percentage Error (MAPE): Measures the accuracy as a percentage.

**Results**

Performance: The final model achieved the following metrics:

Mean Absolute Error: 0.8121476569407605

Mean Absolute Error (MAE): 0.8121476569407605

Mean Squared Error (MSE): 1.1668217550839965

Root Mean Squared Error (RMSE): 1.080195239335925

R-squared (R²) Score: 0.38818821363625333

Mean Absolute Percentage Error (MAPE): 16.847194124061353

**Key Insights**

Genre: Certain genres tend to receive higher ratings.

Director: Movies directed by well-known directors often have higher ratings.

Actors: The presence of popular actors positively influences ratings.

Votes:Movies with a higher number of votes generally have more reliable ratings.

**Key Learnings**

Data Preprocessing: The importance of handling missing data and creating meaningful features to improve model accuracy.

Model Evaluation: Understanding the trade-offs between different performance metrics and selecting the best model based on balanced performance.

Prediction: Gaining practical experience in using a trained model to predict ratings for new movies.

**GitHub Repository**

Explore the complete project and code here:

<https://github.com/sravanthi224/MOVIE-RATING-PREDICTION.git>