Assignment: Predicting Movie Ratings for Indian Films

Introduction:
The purpose of this assignment is to develop a machine learning model that predicts the rating of Indian movies based on various features. The dataset contains information about movie name, release year, duration, genre, number of votes, director, lead actors and the movie's rating. The goal is to build a model that can accurately predict the movie's rating based on the combination of year, duration, genre, director, and lead actors.

Dataset: You will be provided with a dataset of Indian movies which contains the following features:

1. Name: The name of the movie
2. Year: The year of release
3. Duration: The running time of the movie in minutes
4. Genre: The genre of the movie (e.g., action, drama, comedy, etc.)
5. Rating: The average rating of the movie (out of 10)
6. Votes: The number of votes the movie has received
7. Director: The name of the movie director
8. Actor 1: The name of the lead actor
9. Actor 2: The name of the second lead actor
10. Actor 3: The name of the third lead actor

Task:

1. Load the dataset into a pandas dataframe and perform some initial exploratory analysis.
2. Pre-process the data as necessary. This could include handling missing values, encoding categorical variables, and scaling numerical variables.
3. Split the data into a training set and a testing set.
4. Train several machine learning models using the training set, including linear regression, decision trees, random forests, and gradient boosting.
5. Evaluate the models' performance using the testing set and select the best model.

6. Fine-tune the selected model using hyperparameter tuning to improve its performance.
7. Make predictions on a few sample movie combinations to demonstrate the model's ability to predict movie ratings based on the combination of year, duration, genre, director, and lead actors.

Deliverables:

1. Jupyter Notebook with all code, comments, and explanations for each step.
2. A report summarizing the exploratory analysis, pre-processing steps, model selection, hyperparameter tuning, and evaluation results.
3. A brief presentation of the project and results.
4. Use various evaluation metrics (e.g., mean absolute error, mean squared error, R-squared) to assess the performance of the model
5. Compare the performance of different models and choose the best one.