MOVIE RATING PREDICTION

Task 2

By

Praveenkumar S

B.tech Artificial Intelligence and Data Science

Task 2

Movie rating prediction with python

Problem :

Movie rating prediction with Python.

Problem definition :

Given features like release year, votes, duration, director, and actors for a movie, develop a machine learning model to predict the movie's IMDb rating.

Introduction

This project aims to build a predictive model for movie ratings based on movie metadata. By applying machine learning algorithms, the model can provide rating predictions for new movies. The dataset used contains details of over 5000 movies.

Libraries Used

- NumPy
- Pandas
- Matplotlib
- Seaborn
- Scikit-learn

Dataset

The dataset contains movie attributes like title, year of release, duration, director, actors, votes, and actual IMDb rating. There are 5043 movie records with 12 features.

Data Preprocessing

Steps like handling missing values, encoding categorical variables like director and actors using LabelEncoder, and converting data types were performed.

Splitting Data

The preprocessed data was split into training (80%) and testing (20%) sets for model building and evaluation.

Model Training

A Random Forest Regressor model was trained on the training data to predict the IMDb rating by capturing complex relationships between different input features.

Model Testing

The model was tested on the unseen testing dataset. Performance was evaluated using R2 score.

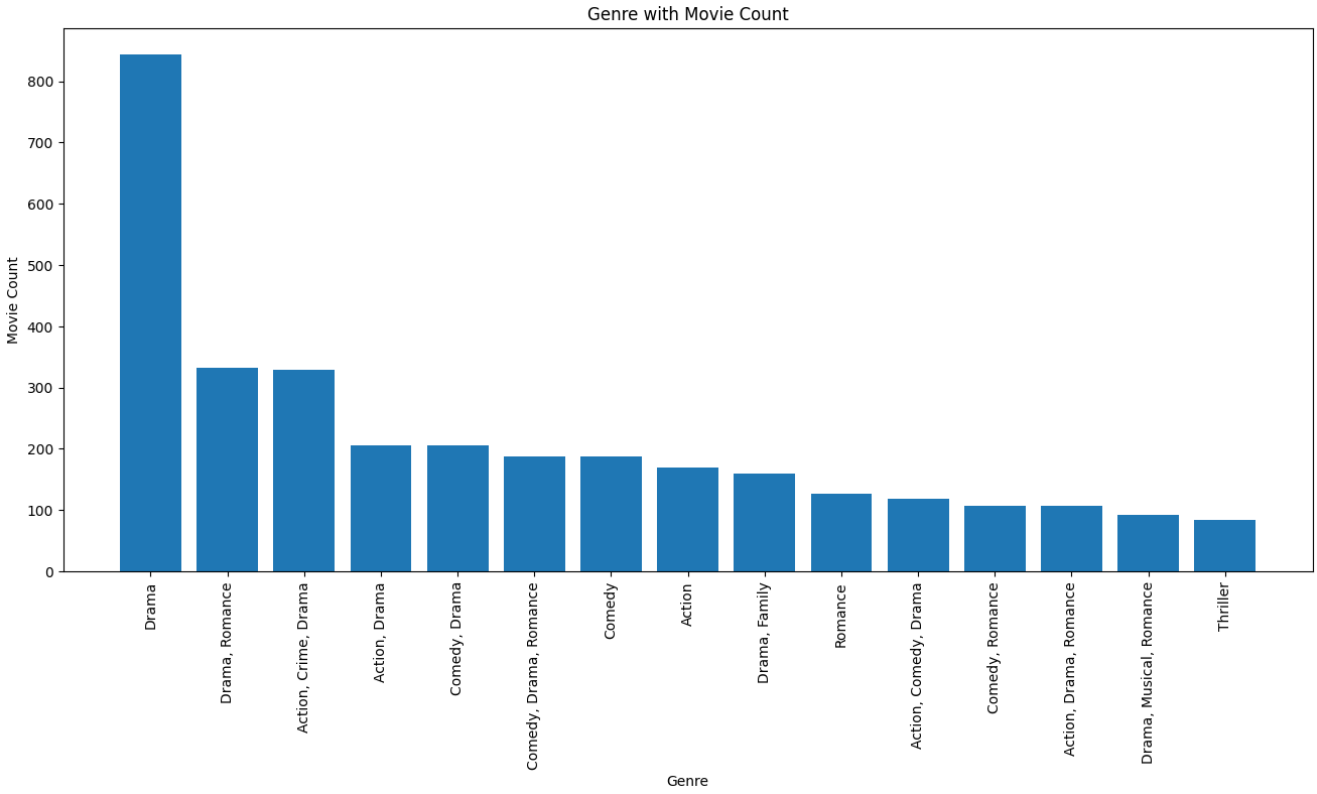
Custom Testing

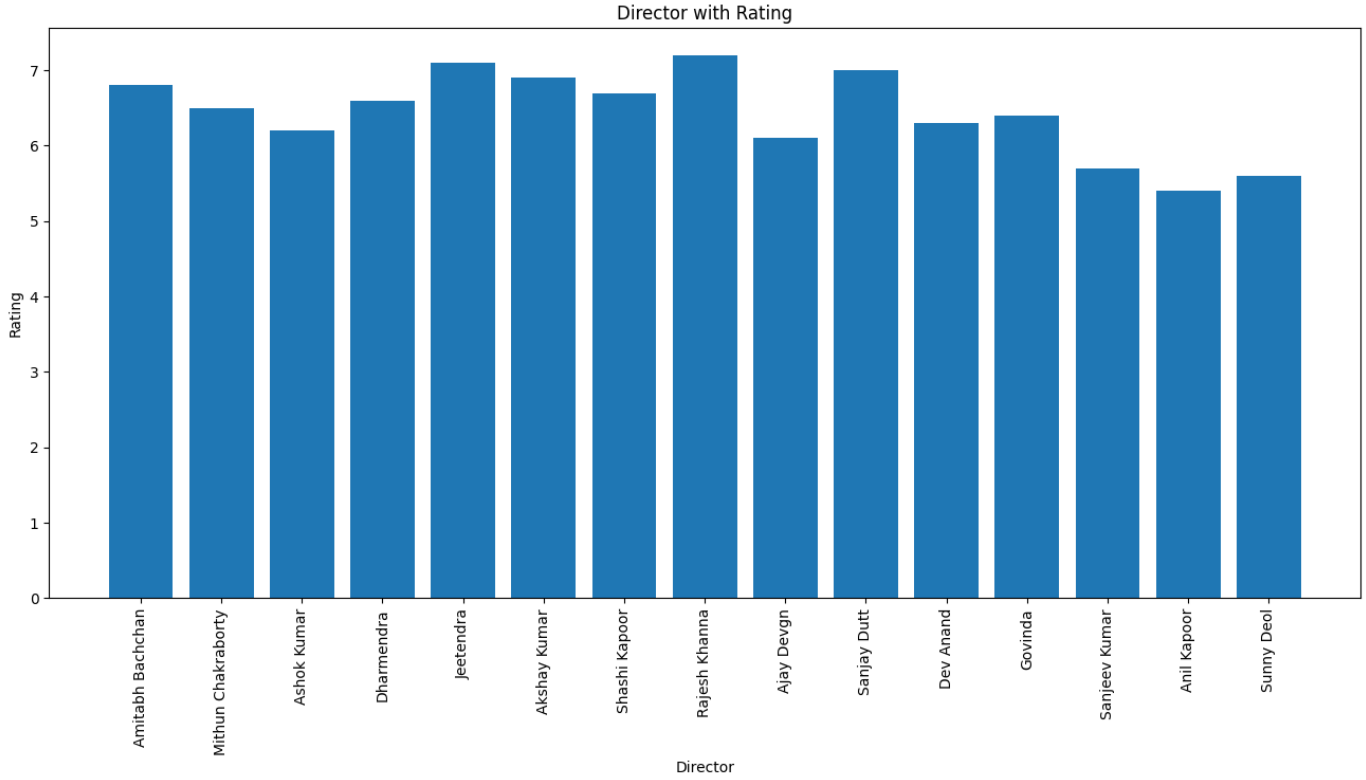
An input sample with details of a hypothetical movie was defined and passed to the trained model to predict its rating. This demonstrates how the model can predict ratings for new movies.

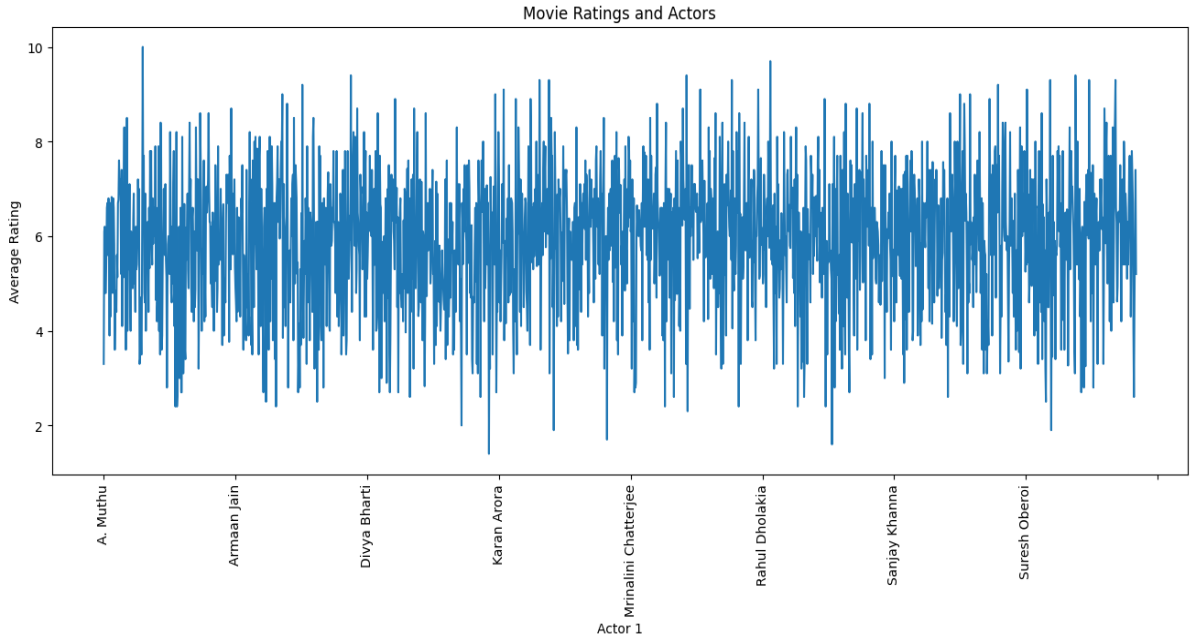
Python Code

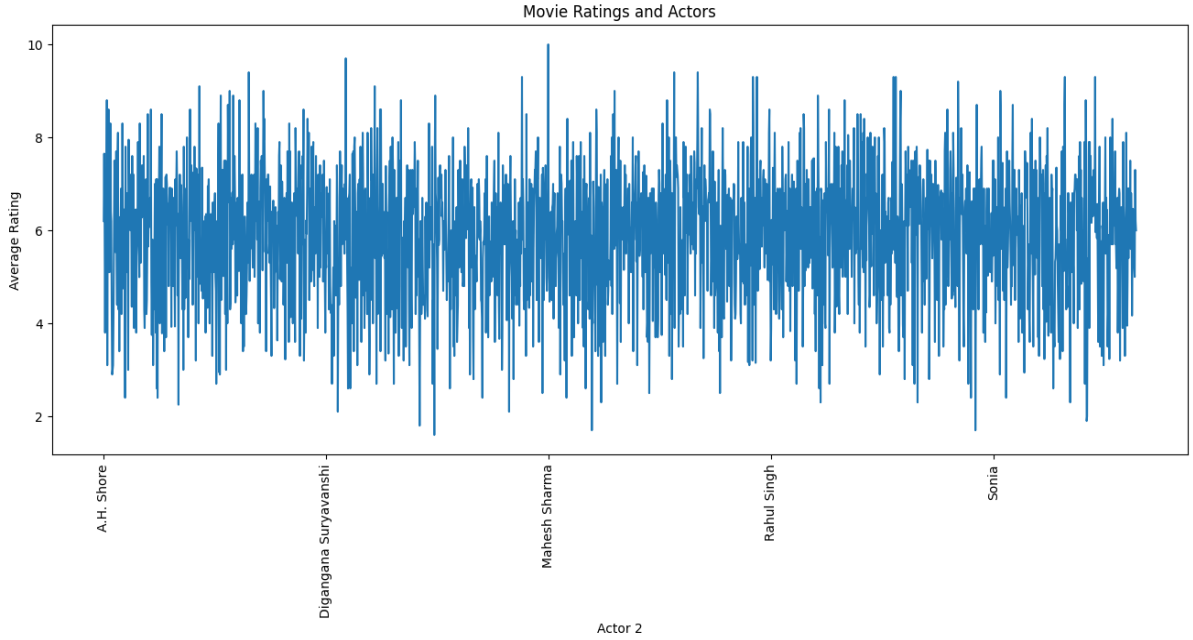
```
year = int(input("Year : "))
Votes = int(input("Votes : "))
Duration = float(input("Duration : "))
Director = float(input("Director : "))
Actor1 = float(input("Actor 1 : "))
Actor2 = float(input("Actor 2 : "))
Actor3 = float(input("Actor 3 : "))
```

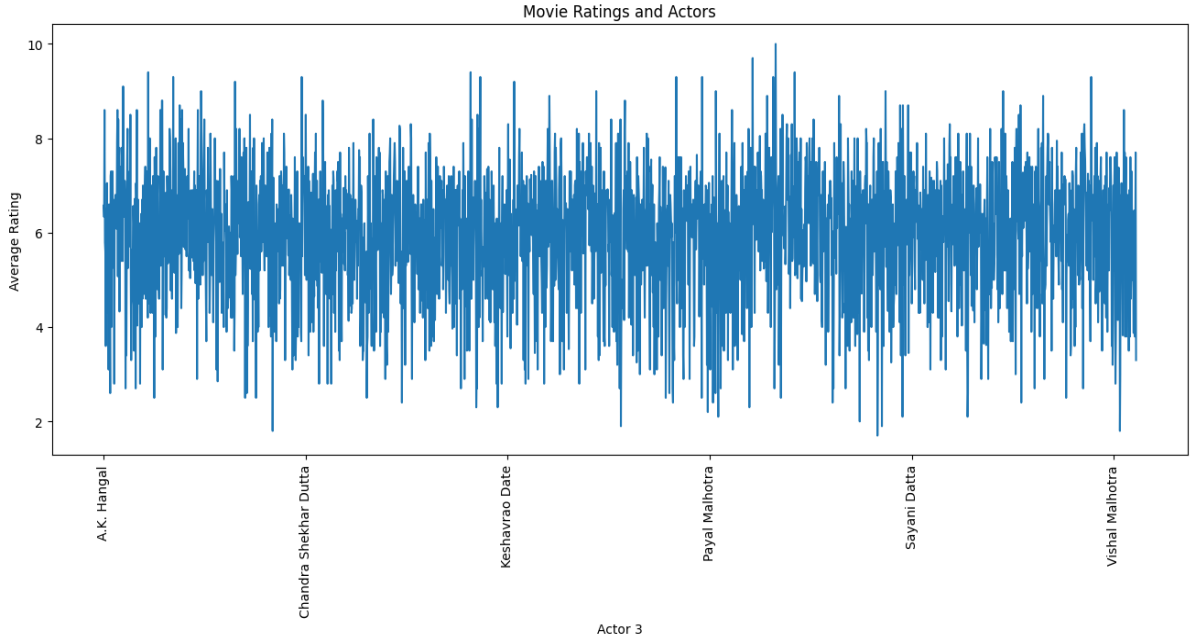
Plots


Genre with Movie Count
Movie Count
0
100
200
300
400
500
600
700
800
Drama
Drama, Romance
Action, Crime, Drama
Action, Drama
Comedy, Drama
Comedy, Drama, Romance
Comedy
Action
Drama, Family
Romance
Action, Comedy, Drama
Comedy, Romance
Action, Drama, Romance
Drama, Musical, Romance
Thriller
Genre


Director with Rating
Rating
0
1
2
3
4
5
6
7
Amitabh Bachchan
Mithun Chakraborty
Ashok Kumar
Dharmendra
Jeetendra
Akshay Kumar
Shashi Kapoor
Rajesh Khanna
Ajay Devgn
Sanjay Dutt
Dev Anand
Govinda
Sanjeev Kumar
Anil Kapoor
Sunny Deol
Director


Movie Ratings and Actors
Average Rating
10
8
6
4
2
A. Muthu
Armaan Jain
Divya Bharti
Karan Arora
Mrinalini Chatterjee
Rahul Dholakia
Sanjay Khanna
Suresh Oberoi
Actor 1


Movie Ratings and Actors
Average Rating
10
8
6
4
2
A.H. Shore
Digangana Suryavanshi
Mahesh Sharma
Rahul Singh
Sonia
Actor 2


Movie Ratings and Actors
Average Rating
10
8
6
4
2
A.K. Hangal
Chandra Shekhar Dutta
Keshavrao Date
Payal Malhotra
Sayani Datta
Vishal Malhotra
Actor 3

Output Sample

```
Python Code

>> Year          : 2015
>> Votes         : 250
>> Duration   : 120
>> Director    : 9.2
>> Actor 1     : 8.1
>> Actor 2     : 9.1
>> Actor 3     : 9


Random Forest Regressor Model Performance:
R2 score  :  0.2659253618245764
>> Predicted rating  :  6.125
```

References

- Google
- Youtube
- Kaggle

Conclusion

The random forest regression model effectively predicts movie ratings based on provided metadata. This model can be integrated into movie recommendation systems or used by movie producers for quality checks.