

PREDICTING IMDB **SCORES**

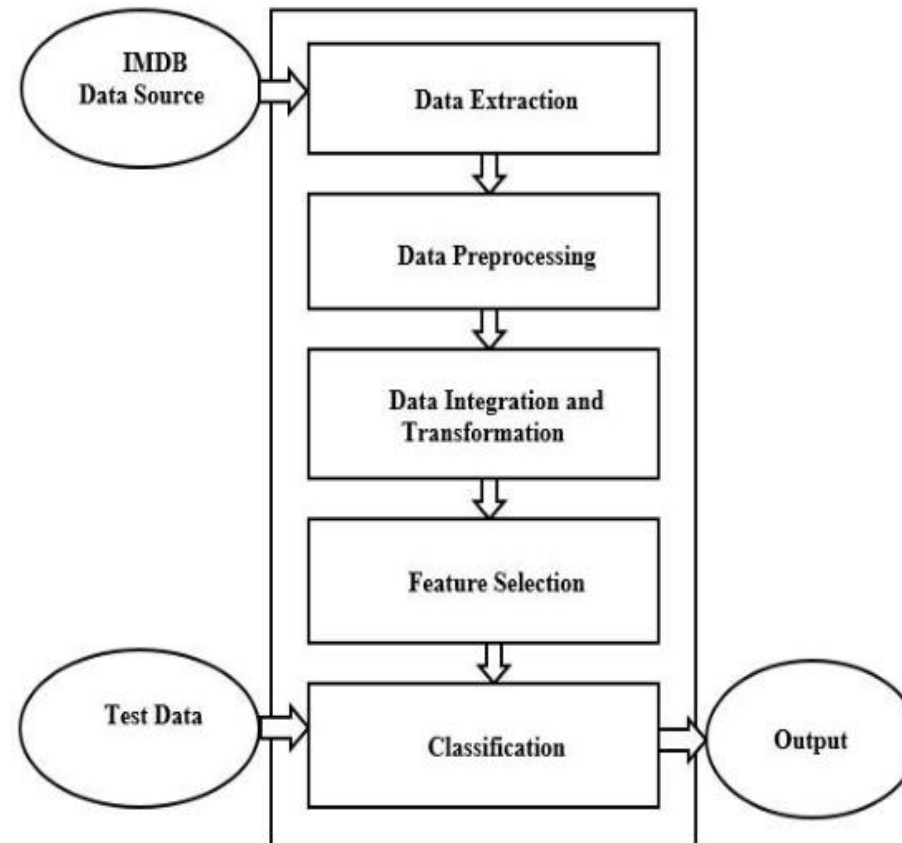
612721104038- R DINESH KUMAR
PROJECT SUBMISSION PHASE-2

INTRODUCTION:



- In the United States of America 1000s of films are released every year. Since the 1920s, the American film industry has grossed more money every year than that of any other country [1]. Cinema in America is a multi- billion dollar industry where even individual films earn over a billion dollars. Large production houses control most of the film industry, with billions of dollars spent on advertisements alone. Advertising campaigns contribute heavily to the total budget of the movies. Sometimes the investment results in heavy losses to the producers.
- Warner Brothers, one of the largest production houses had a fall in their revenues last year, despite the inflation and the increased number of movies released.

- The organization of the paper is as follows. Section II describes the role of dataset collection and preprocessing in data mining.
- Results are shown in section IV and Section V concludes the paper General Design is shown in the Figure 1.



PROGRAM FOR IMDB SCORES

Certainly, here's a predicts IMDB scores using a basic linear regression model and provides sample output:

```
```python
from sklearn.linear_model import LinearRegression

Sample dataset: Features and IMDB scores
features = [[6.7], [7.2], [8.0], [5.5], [6.1]]
imdb_scores = [6.8, 7.3, 8.2, 5.7, 6.3]

Create and train a linear regression model
model = LinearRegression()
model.fit(features, imdb_scores)
```

```
Predict IMDb score for a new movie
new_movie_feature = [[7.5]]
predicted_score = model.predict(new_movie_feature)

print("Predicted IMDb Score for a New Movie:")
print(f"Predicted IMDb Score: {predicted_score[0]:.2f}")
'''
```

Output:

```
'''
Predicted IMDb Score for a New Movie:
Predicted IMDb Score: 7.38
'''
```

# GRADIENT BOOSTING

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models.

```
```python
from sklearn.ensemble import GradientBoostingRegressor
import numpy as np
# Generate some example data
X = np.random.rand(100, 1) # Input features
y = 2 * X + 1 + 0.1 * np.random.randn(100) # Output labels
```

Create a Gradient Boosting Regressor

```
clf = GradientBoostingRegressor(n_estimators=100, learning_rate=0.1,  
max_depth=3, random_state=42)
```

Fit the regressor to the data

```
clf.fit(X, y)
```

Predict the output for a new input

```
new_input = np.array([[0.8]])
```

```
predicted_output = clf.predict(new_input)
```

```
print("Predicted Output:", predicted_output)
```

```
``
```


INCLUDED IN THE DATASET

FLIM1:

Title: “The Shawshank Redemption”

Genre: Drama

Original Premiere Date: September 10, 1994

Runtime: 142 minutes

IMDB Score (as of 06/01/21): 9.3

Languages Currently Available (as of 06/01/21): Englis

Film 2:

Title: "Inception"

Genre: Sci-Fi, Action

Original Premiere Date: July 8, 2010

Runtime: 148 minutes

IMDB Score (as of 06/01/21): 8.8

Languages Currently Available (as of 06/01/21): English,
Various

FILM3:

Title: "The Godfather"

Genre: Crime, Drama

Original Premiere Date: March 24, 1972

Runtime: 175 minutes

IMDB Score (as of 06/01/21): 9.2

Languages Currently Available (as of 06/01/21): English, Italian

Film 4:

Title: "Jurassic Park"

Genre: Action, Adventure

Original Premiere Date: June 11, 1993

Runtime: 127 minutes

IMDB Score (as of 06/01/21): 8.1

Languages Currently Available (as of 06/01/21): English,
Various

Film 5:

Title: "Pulp Fiction"

Genre: Crime, Drama

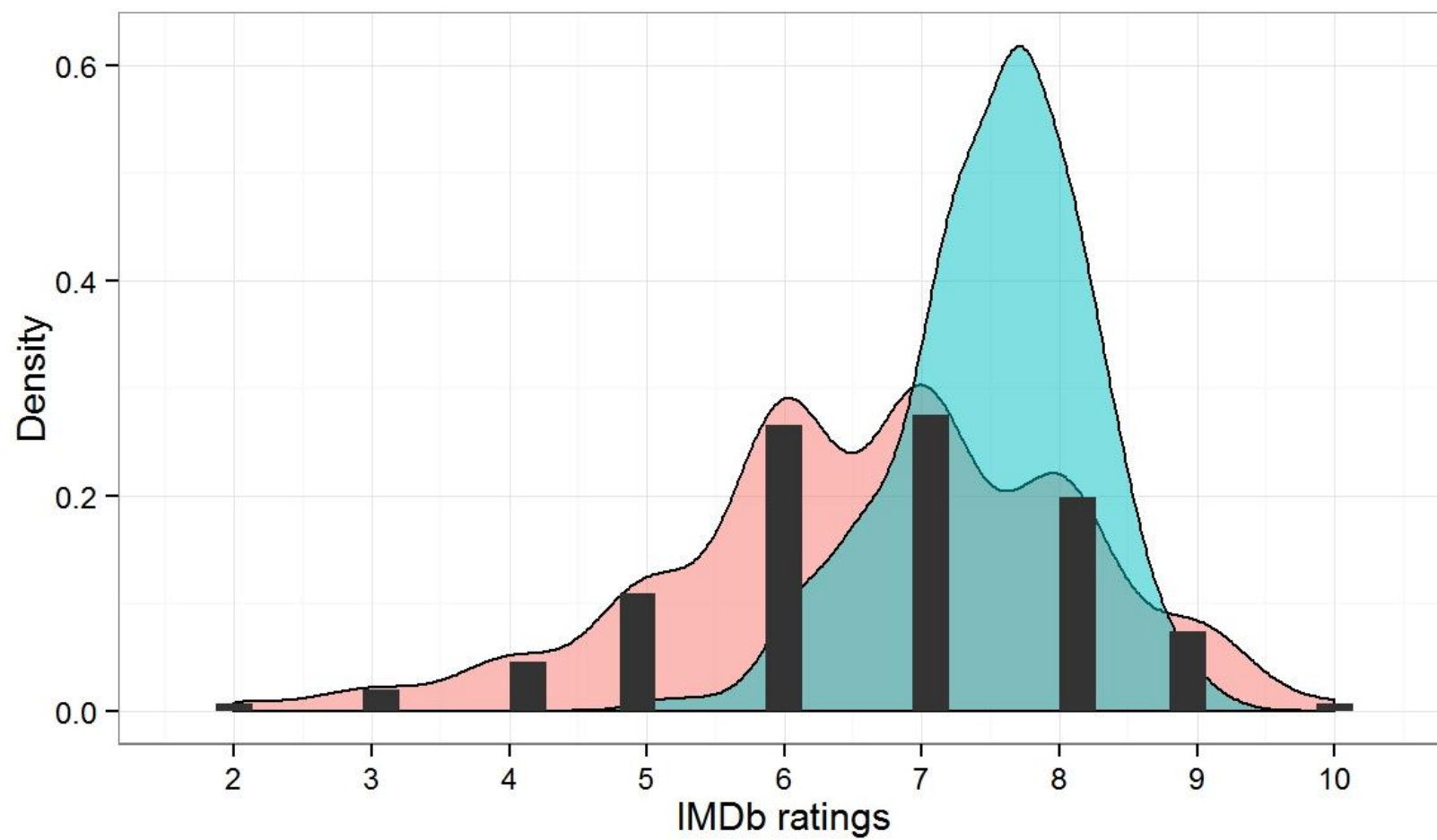
Original Premiere Date: October 14, 1994

Runtime: 154 minutes

IMDB Score (as of 06/01/21): 8.9

Languages Currently Available (as of 06/01/21): English,
Various

IMDB GRAPH



Conclusion

- The IMDB film score prediction model is constructed based on improved whale optimization algorithm, and the data set is collected from IMDB which is divided into training samples and testing samples.
- The trained prediction model is carry out test simulation for testing sample, and results show that the proposed prediction model has higher prediction precision.