BVRIT HYDERABAD

College of Engineering for Women

MOVIE CRITIC

Team No : 9
Team Members:
M. Sandhya - 20wh1a1243
P. Harini - 20wh1a1244
E. Meenakshi Supriya - 20wh1a1245
G. Indhu - 20wh1a1246
Madeeha Qamar - 20wh1a1247

AGENDA

- Problem statement
- Python Packages used
- Types of Algorithms
- Output and Graph
- Comparison table
- Execute the Code

VISHNU
UNIVERSAL LEARNING

Problem Statement

- In this dataset, you are provided with 7398 movies. Movies are labeled with id. Data points include cast, crew, plot keywords, budget, posters, release dates, languages, production companies, and countries. You are predicting the worldwide revenue for 4398 movies in the test file. Note - Many movies are remade over the years, therefore it may seem like multiple instances of a movie appear in the data, however they are different and should be considered separate movies. In addition, some movies may share a title, but be entirely unrelated. E.g. The Karate Kid (id: 5266) was released in 1986, while a remake (id: 1987) was released in 2010.

VISHNU
UNIVERSAL LEARNING

Python Packages used

- pandas
- numpy
- scikit-learn
- matplotlib
- seaborn
- datetime
- collections
- ast

VISHNU
UNIVERSAL LEARNING

Algorithm

- Multiple Linear Regression
- Random Forest
- Ridge Regression
- Lasso Regression
- Extreme Gradient Boosting(XGBoost)

VISHNU
UNIVERSAL LEARNING

Multiple Linear Regression

- Linear Regression is a machine learning algorithm based on supervised learning.
- It performs a regression task.
- It is a statistical method that is used for predictive analysis.
- Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables.

VISHNU
UNIVERSAL LEARNING

Random Forest

- Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.
- Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.
- Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.
- The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

VISHNU
UNIVERSAL LEARNING

Ridge Regression

- Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.

VISHNU
UNIVERSAL LEARNING

Lasso Regression

- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator
- It is also called as L1 regularization.

VISHNU
UNIVERSAL LEARNING

XGBoost

- XGBoost stands for “Extreme Gradient Boosting”
- XGBoost is an optimized distributed gradient boosting library designed for efficient and scalable training of machine learning models.
- It is an ensemble learning method that combines the predictions of multiple weak models to produce a stronger prediction.
- It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems.

VISHNU
UNIVERSAL LEARNING

Output

id	revenue
3001	636631.1606295245318
3002	3597515.2157505536293
3003	971851.162917065463
3004	13049602.177637019745
3005	17157061.606449148407
3006	428303.86532665085312
3007	632940.85468068075113
3008	11293729.218483906074
3009	26696884.153116011645
3010	346070740.35508292814

Linear regression

VISHNU
UNIVERSAL LEARNING

Output

id	revenue
3001	186946.02166424189247
3002	909296.72912333701595
3003	932730.65962322913776
3004	11464618.527065978673
3005	1196897.1219268795201
3006	871589.5577037696444
3007	548864.2473859212043
3008	33145495.880680322032
3009	27089572.693474583539
3010	151628027.57217869734

Random Forest

VISHNU
UNIVERSAL LEARNING

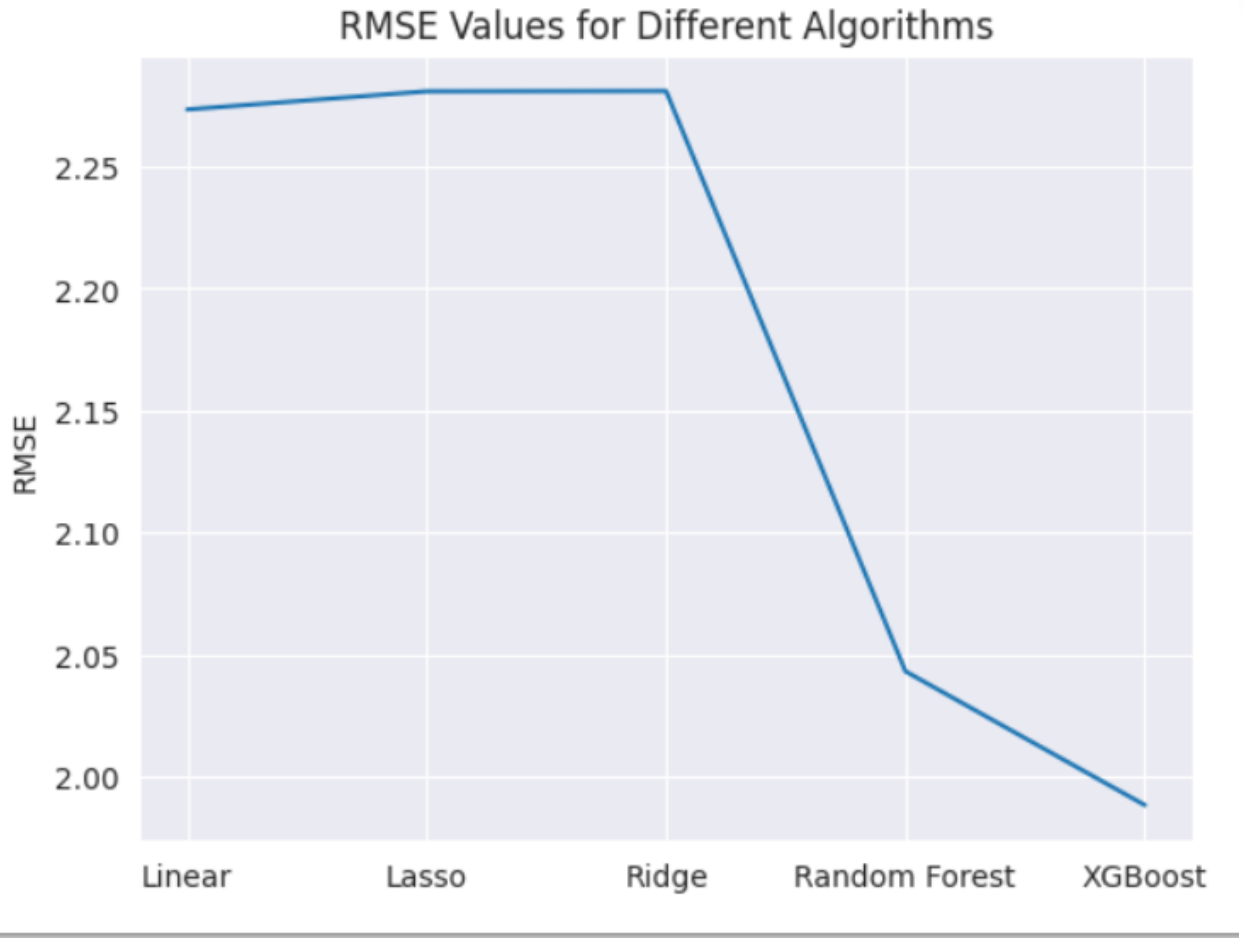
Output

id	revenue
3001	593741.4373866120468
3002	1687587.1218405323546
3003	1112642.0180739149133
3004	18149667.000749076815
3005	2177403.5283906561785
3006	536798.214941865891
3007	356164.11882848253111
3008	16346426.967385990254
3009	21819474.976510670012
3010	190044884.84351855755

XGBoost

VISHNU
UNIVERSAL LEARNING

Graph



VISHNU
UNIVERSAL LEARNING

Comparision Table

	Model	RMSE
0	Linear Regression	2.273472
1	Lasso Regression	2.280914
2	Ridge Regression	2.281044
3	Random Forest	2.043063
4	XGBoost	1.988186

Execute Code

Budget in 1 million:

16.45457

Does the movie has tagline?

1

is the production country United States of America?:

1

Number of cast members:

24

Number of male cast members:

10

Number of crew members:

72

No of female crew members

0

No of male crew members:

13

No.of members in sound department:

10

No.of members in art department:

12

No.of members in editing department:

2

Enter the ratio of budget and release year:

0.008658

Predict

Predicted revenue: 31,451,824.000

VISHNU
UNIVERSAL LEARNING

THANK YOU