

Second-Hand Car Price Prediction

Simge Çınar

Necati Furkan Çolak

Emre Can Şen

Serkan Uygun



Contents

- 1 Aim & Motivation
- 2 Data Preprocessing
- 3 Prediction Models
- 4 Model Comparison
- 5 Conclusion

Motivation

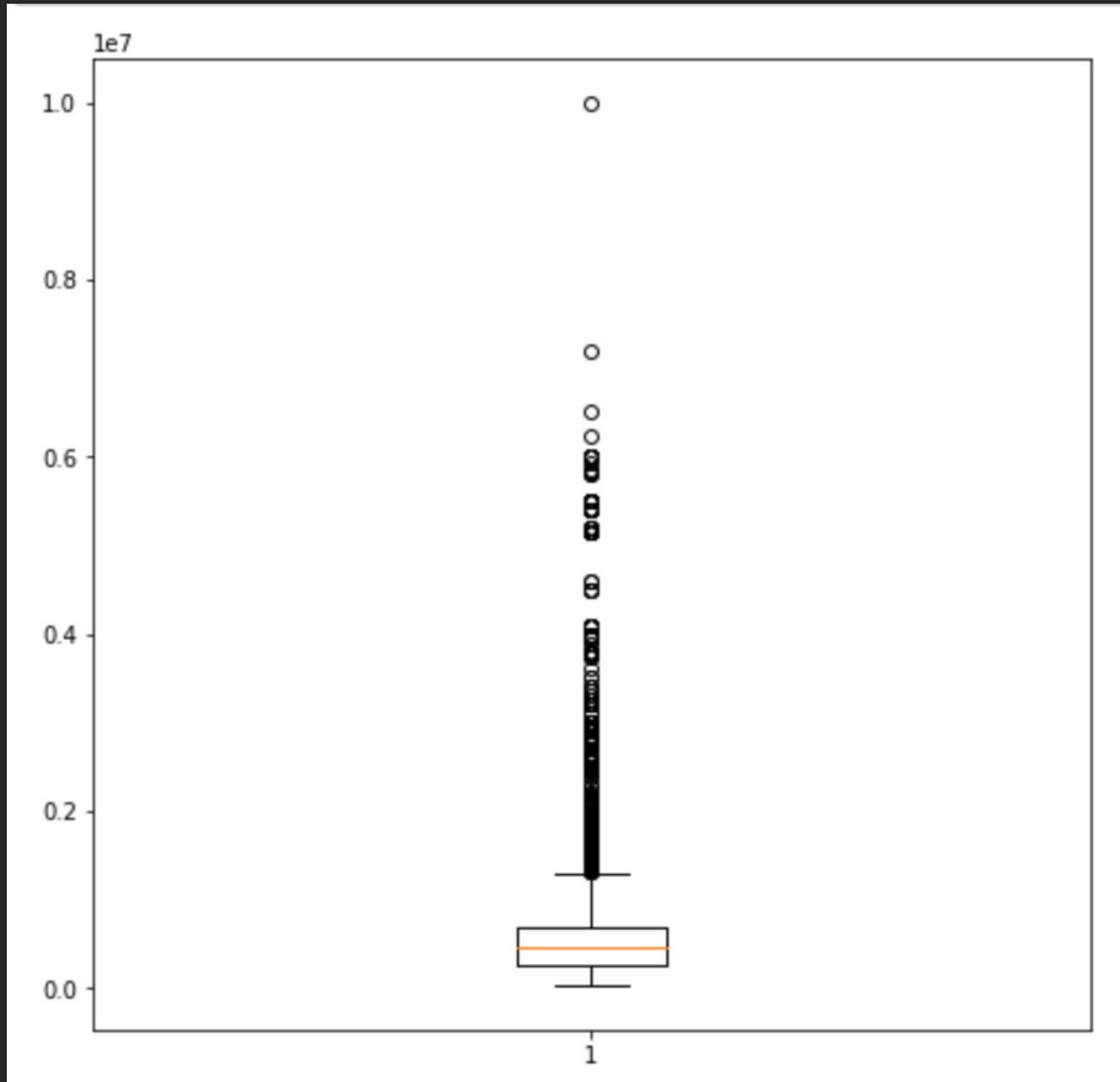
Second-Hand Car
price prediction is
important

Prediction with
machine learning

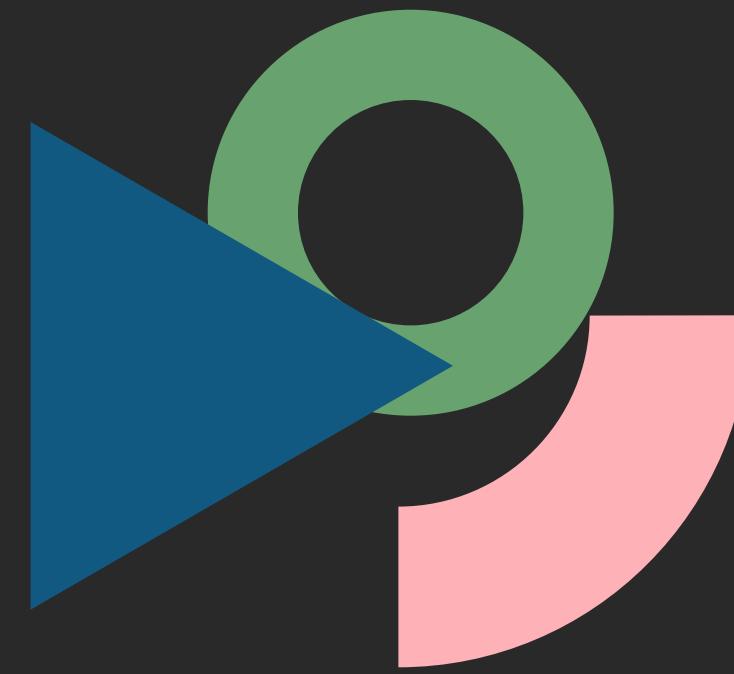
Different ML algorithms

Increase profit

Data Analysis

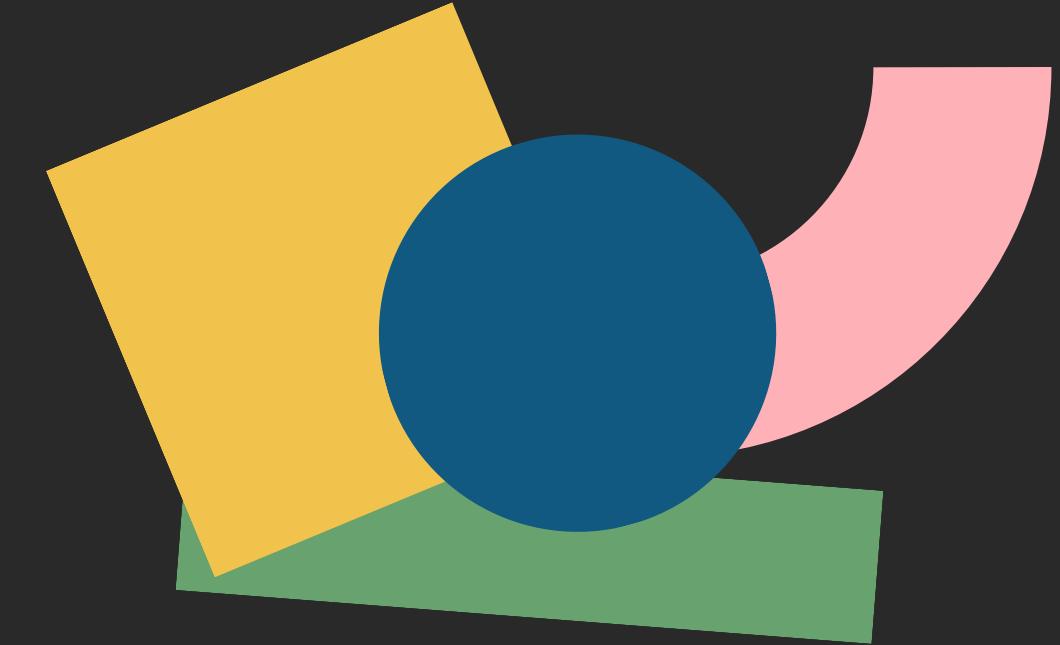


- 8127 rows, 13 columns
- Different feature like year, seats, km_driven, transmission...
- Many outliers in the selling price values (see the boxplot)



Data Transformation

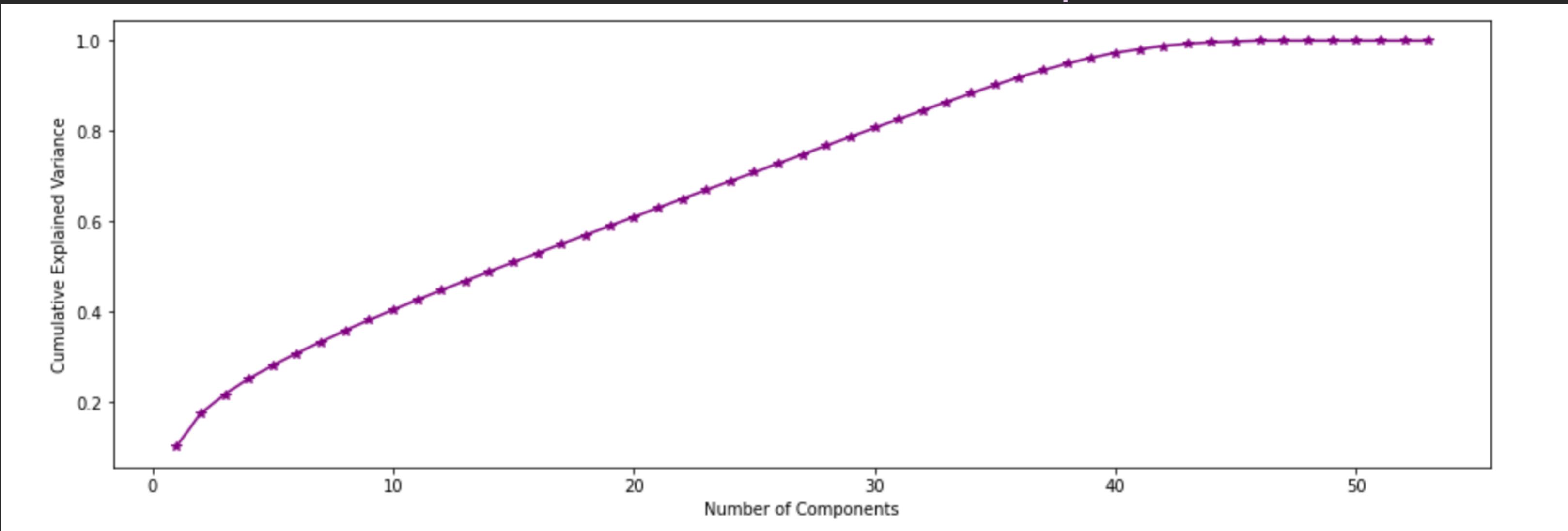
- Data is scaled using Standard Scaler
- PCA is applied after One Hot Encoding since data consists of multiple classes.



Old dimension: 53

New dimension: 39

Explained Variance: 95%



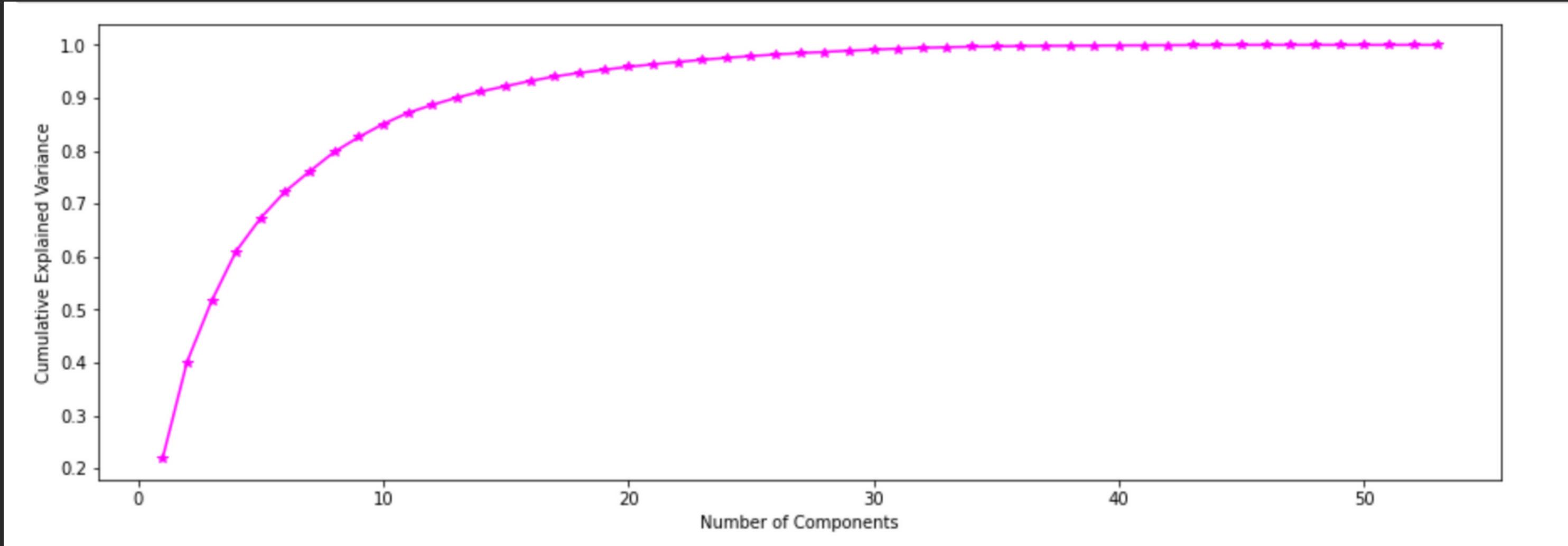
Data Transformation



Old dimension: 53

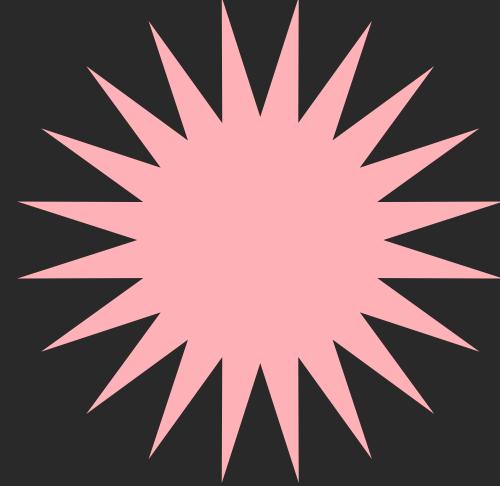
New dimension: 19

Explained Variance: 95%

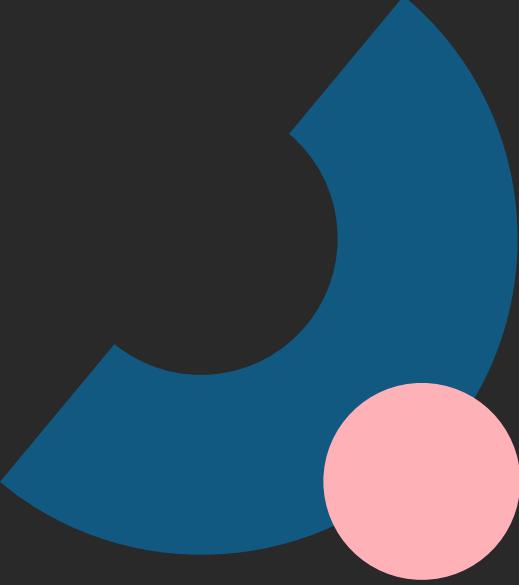


- Min-Max Scaler was also tried
- It yields the same scores in regressions and worse scores in the rest of the models
- Hence Standard Scaler was chosen.





Methods



1- Linear Regression

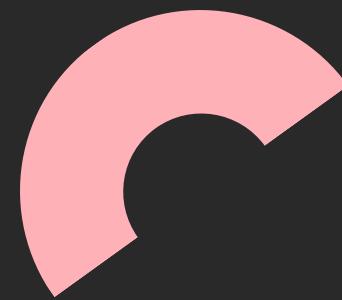
**2- Lasso/Ridge
Regression**

3- Elastic Net

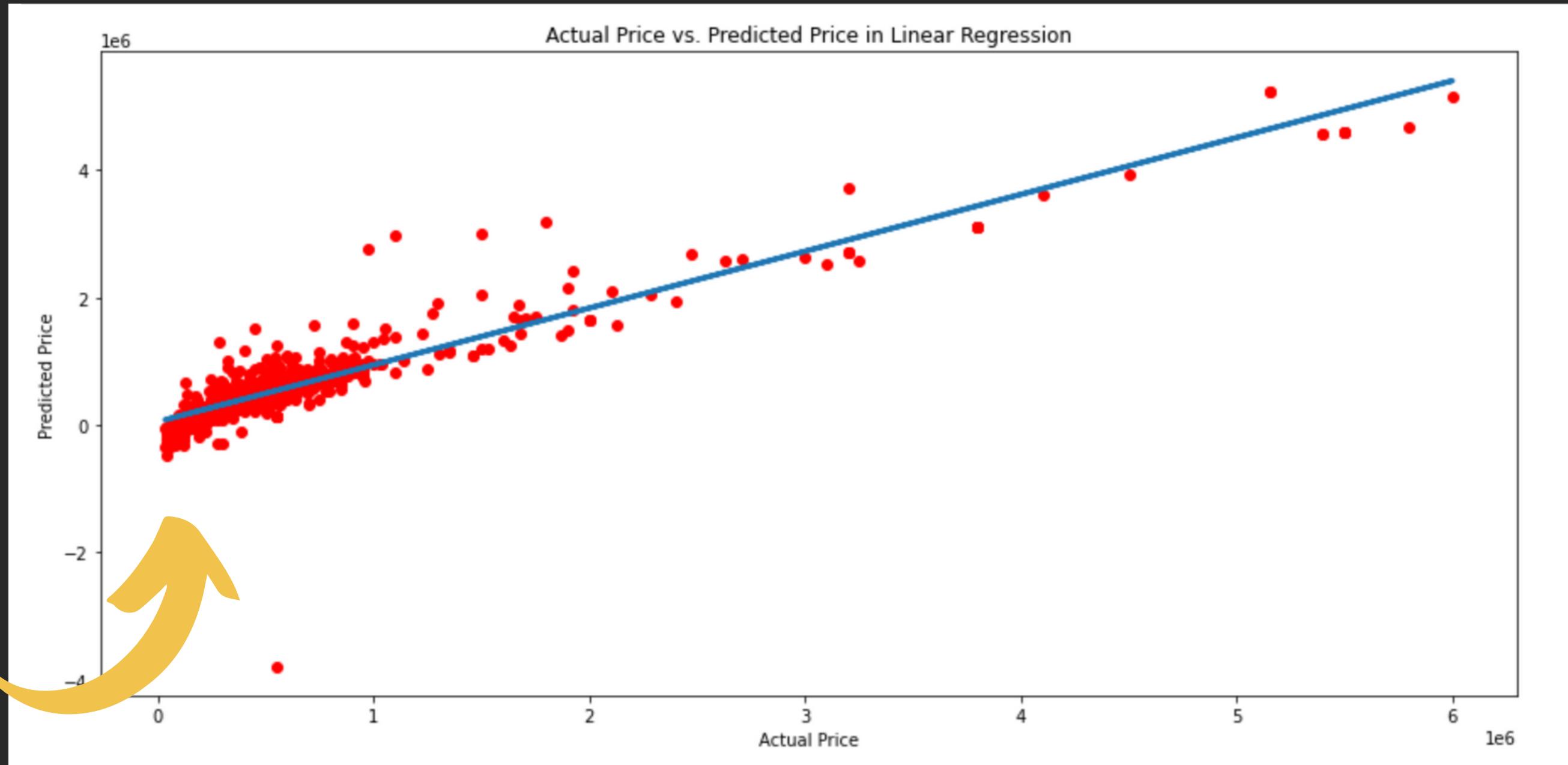
**4- Random Forest
Regression**

**5- Support Vector
Regression**

6- Neural Network



Linear Regression

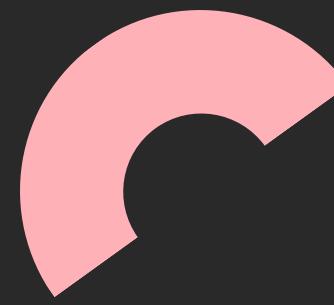


SCORES ON VALIDATION DATA

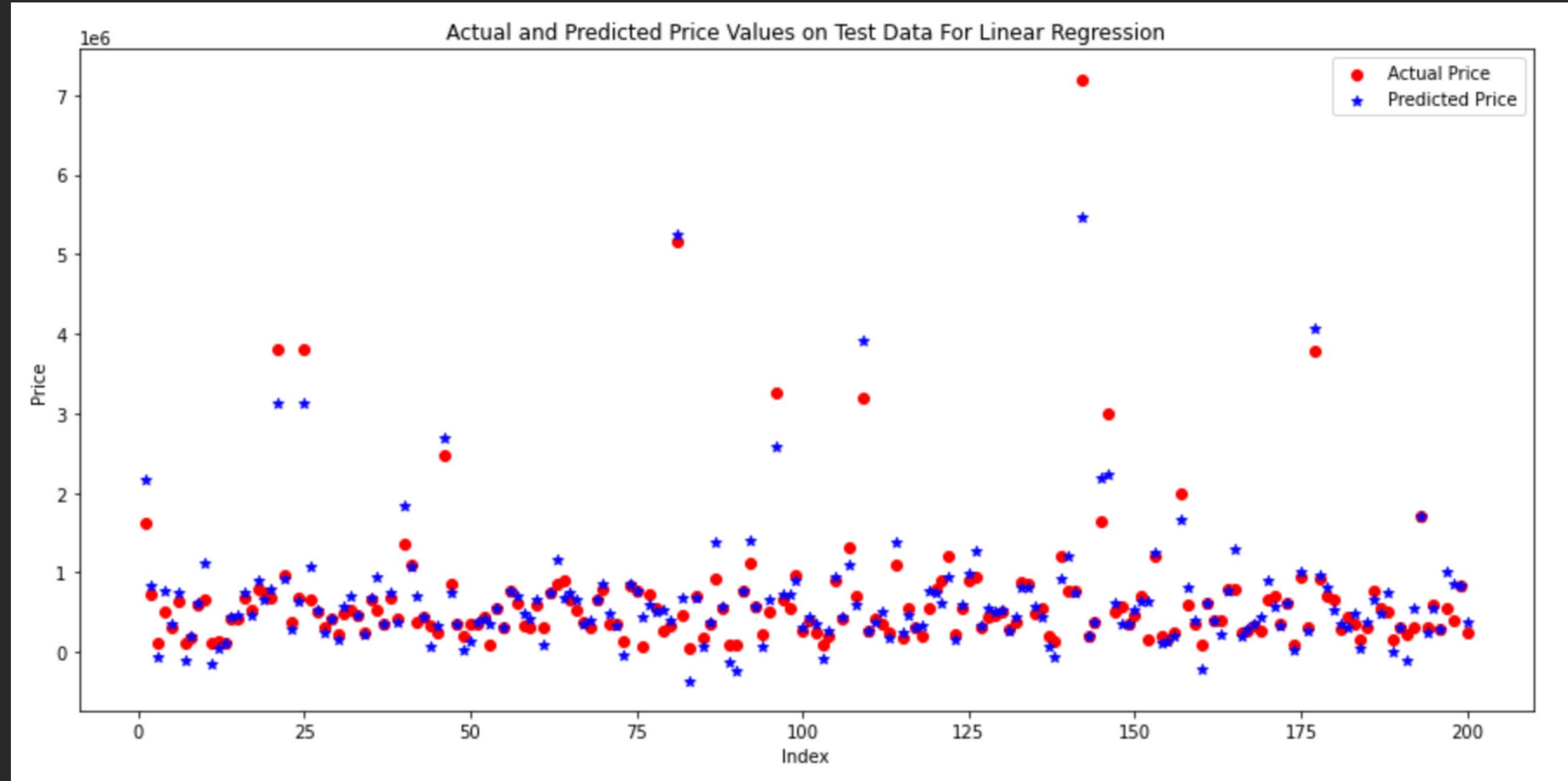
RMSE: 296542.0347

MAPE: 0.5023

R²Score: 0.8553



Linear Regression



SCORES ON TEST DATA

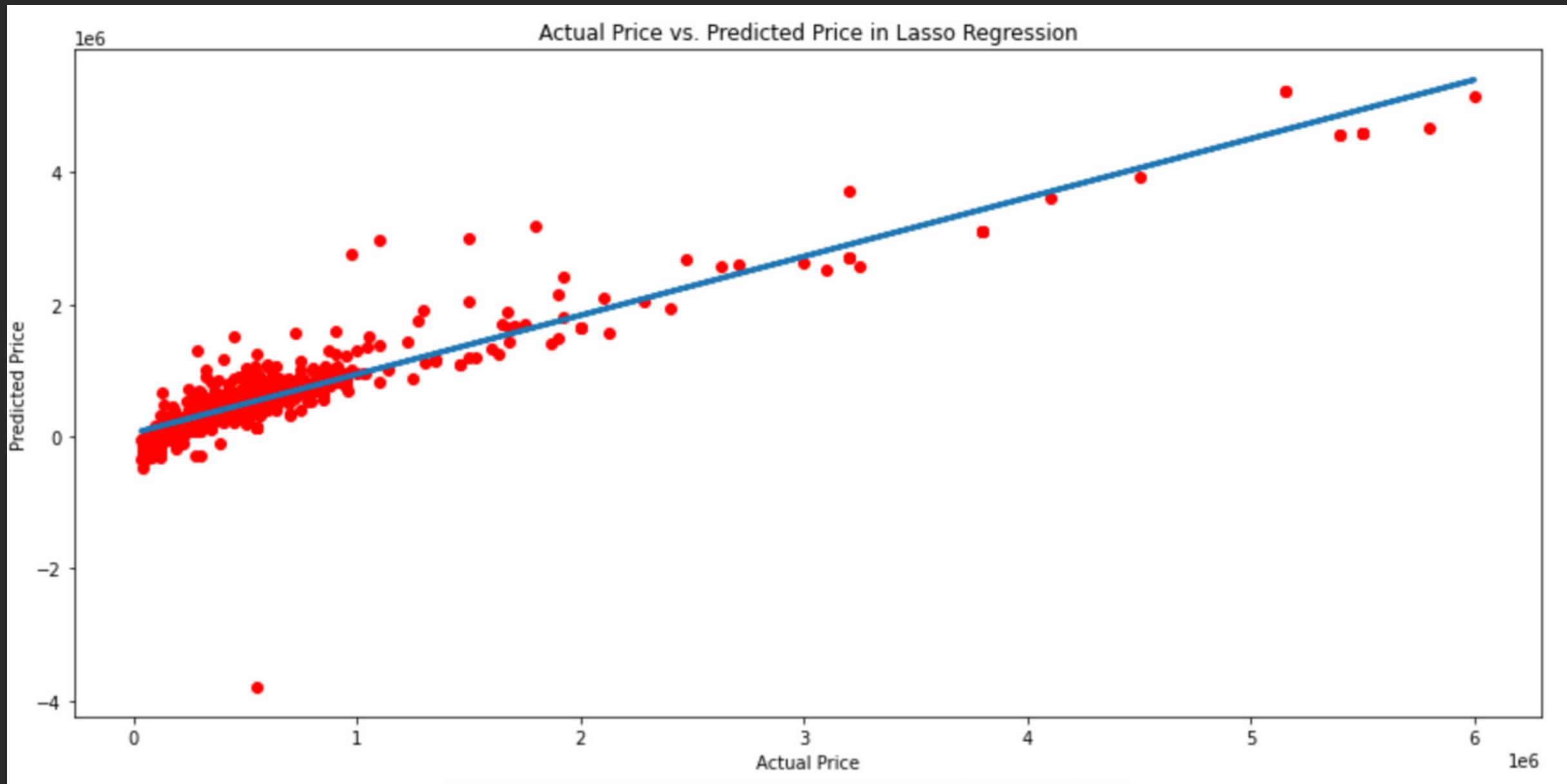
RMSE: 279506.6463

MAPE: 0.4323

R²Score: 0.8748

Lasso Regression

LassoCV yields: alpha = 0.99



SCORES ON VALIDATION DATA

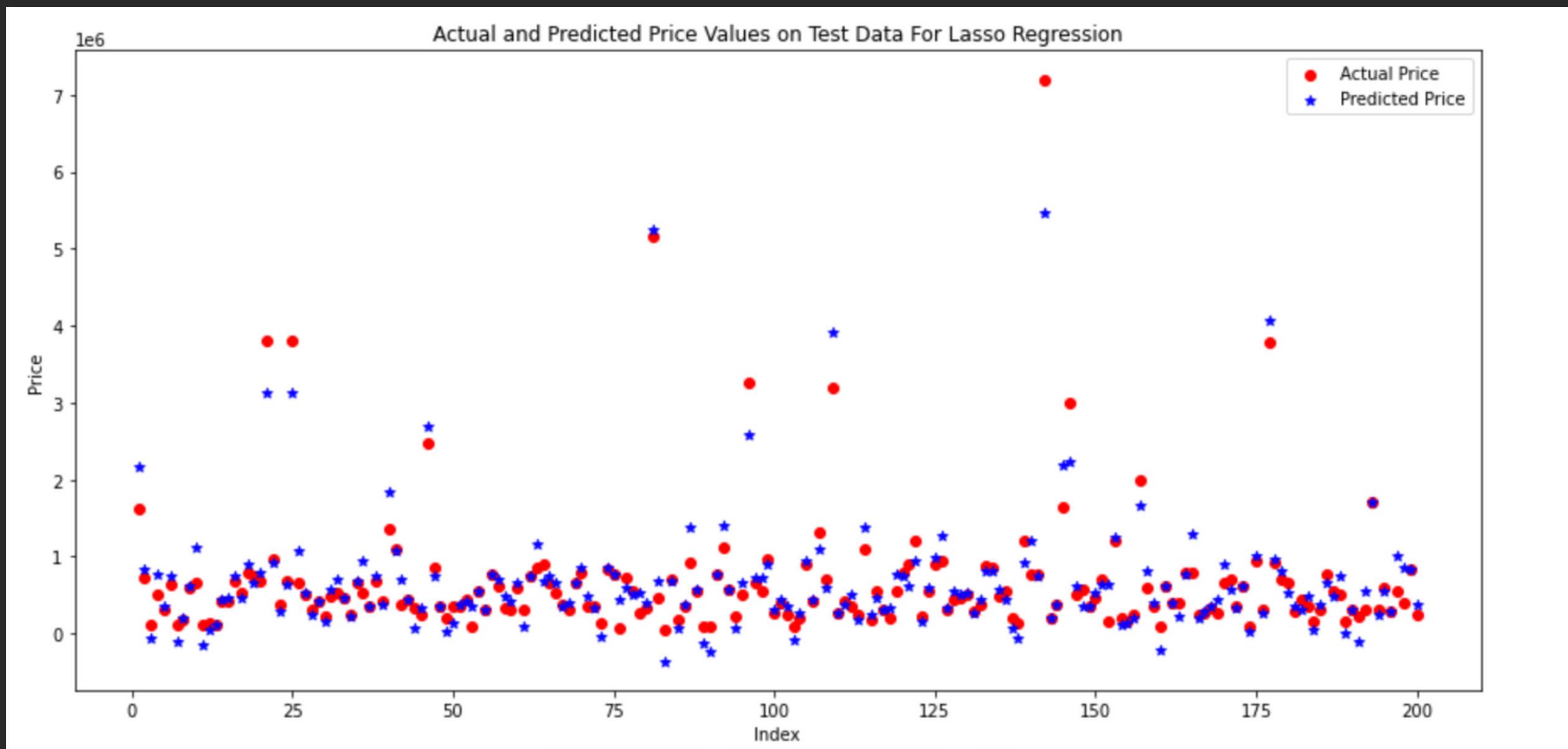
RMSE: 296541.2904

MAPE: 0.5023

R²Score: 0.8553

Lasso Regression

LassoCV yields: alpha = 0.99



SCORES ON TEST DATA

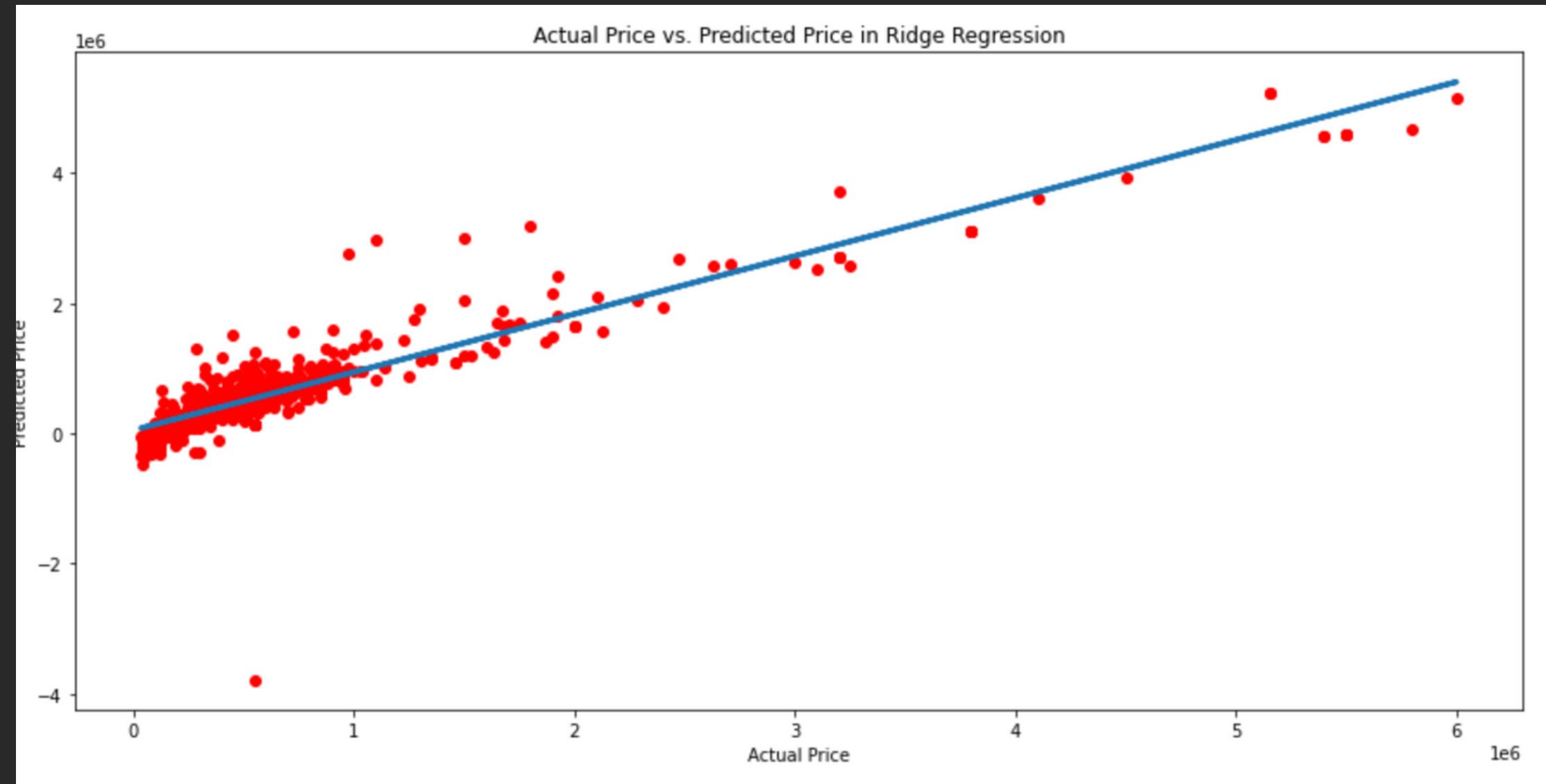
RMSE: 279506.3466

MAPE: 0.4323

R² Score: 0.8748

Ridge Regression

RidgeCV yields: alpha = 0.99



SCORES ON VALIDATION DATA

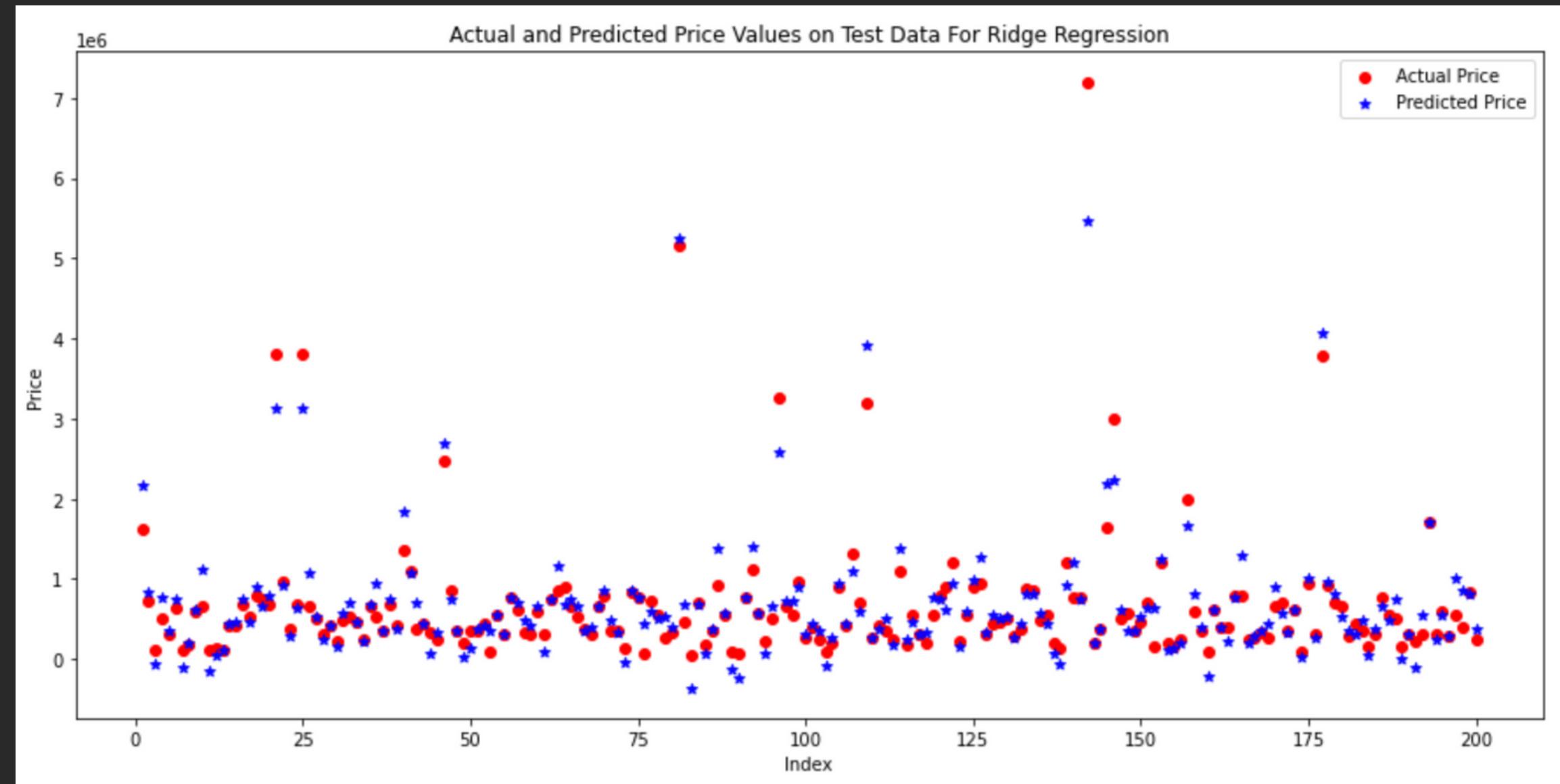
RMSE: 296533.3638

MAPE: 0.5022

R^2 Score: 0.8553

Ridge Regression

LassoCV yields: alpha = 0.99



SCORES ON TEST DATA

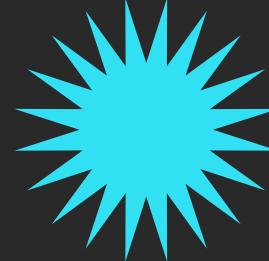
RMSE: 279502.7138

MAPE: 0.4322

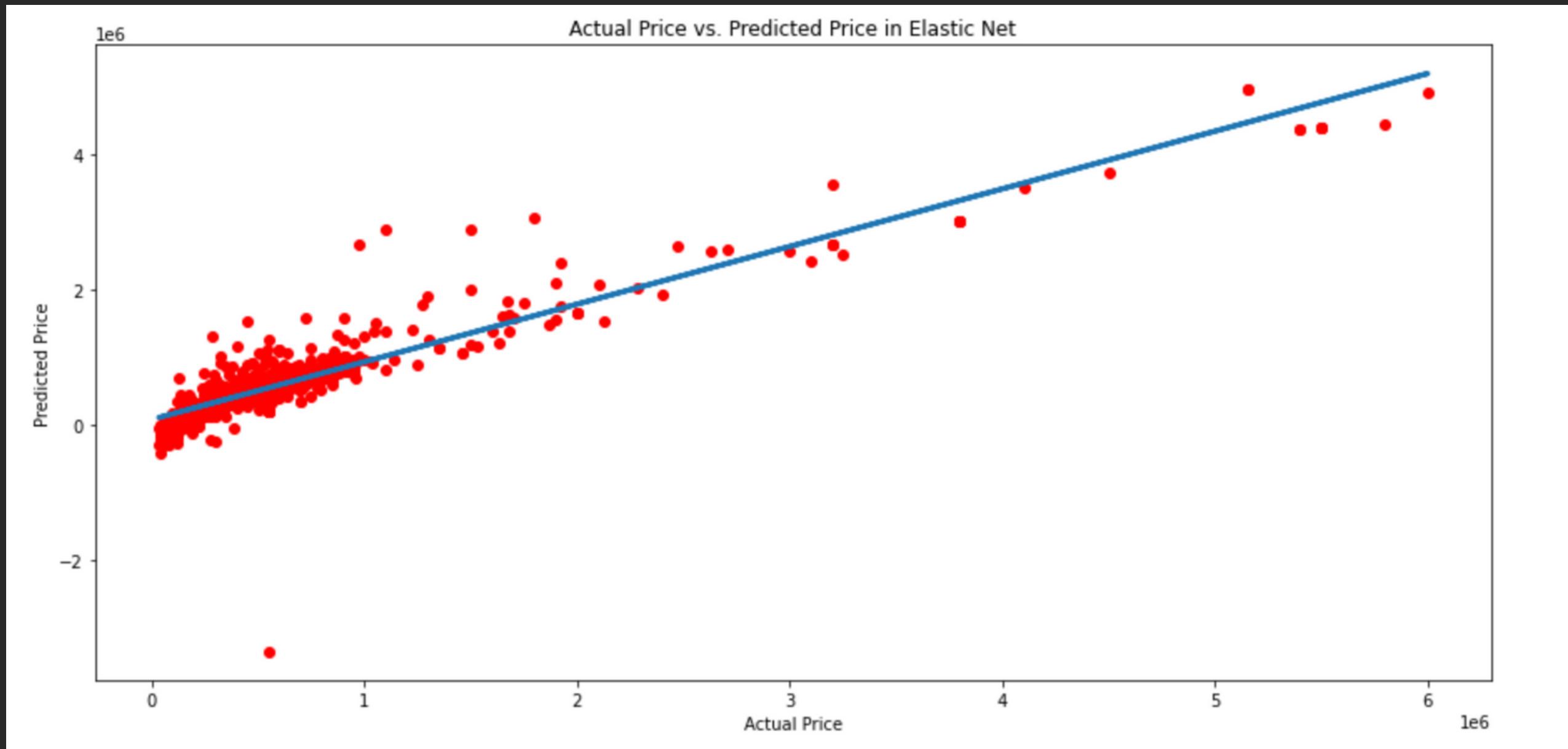
R²Score: 0.8748



ElasticNet Regression



ElasticNetCV yields: alpha = 0.97, l1_ratio = 0.9



SCORES ON VALIDATION DATA

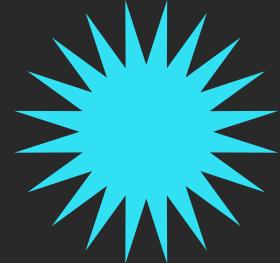
RMSE: 293730.5070

MAPE: 0.4777

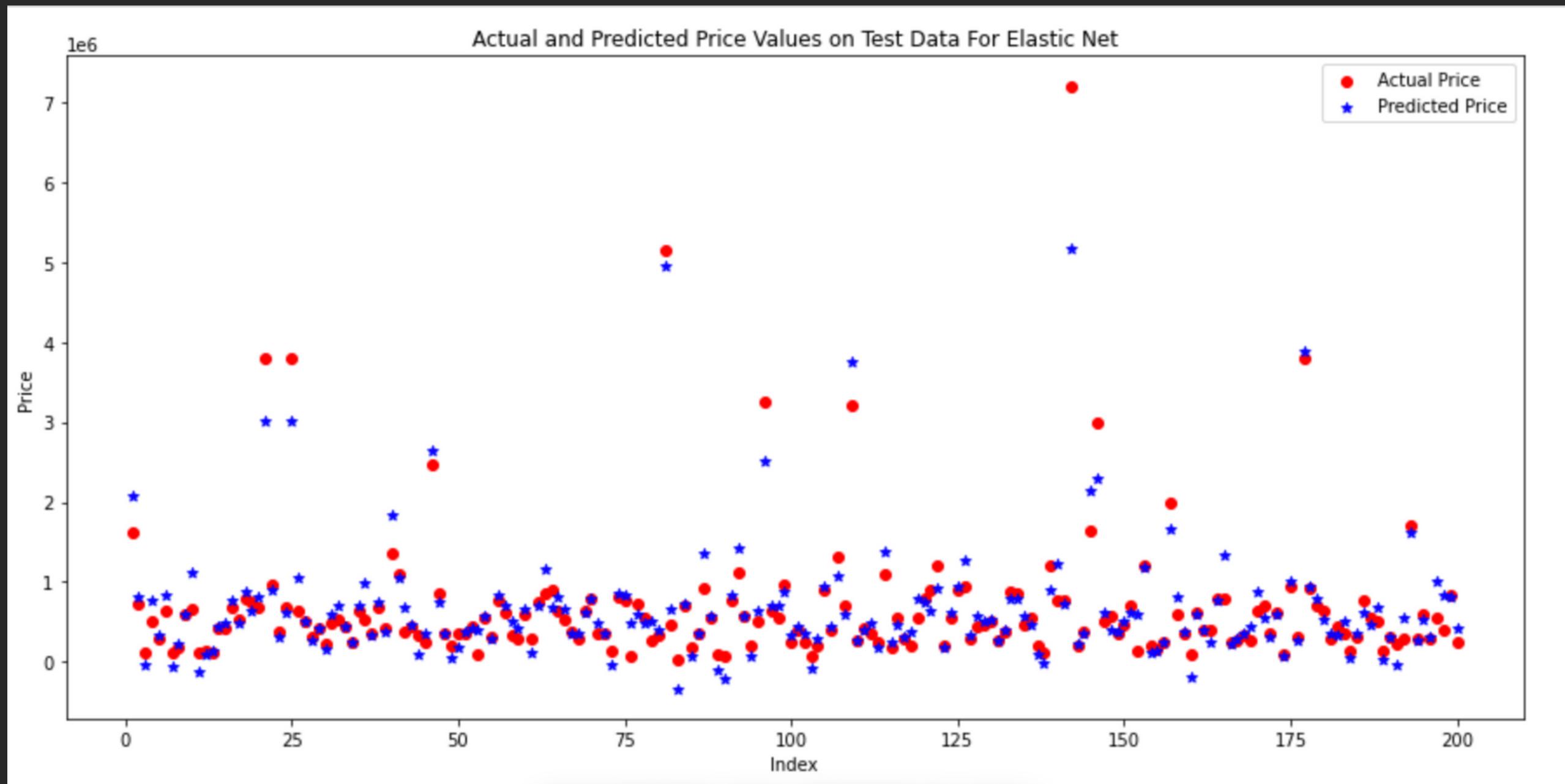
R²Score: 0.8580



ElasticNet Regression



ElasticNetCV yields: alpha = 0.97, l1_ratio = 0.9



SCORES ON TEST DATA

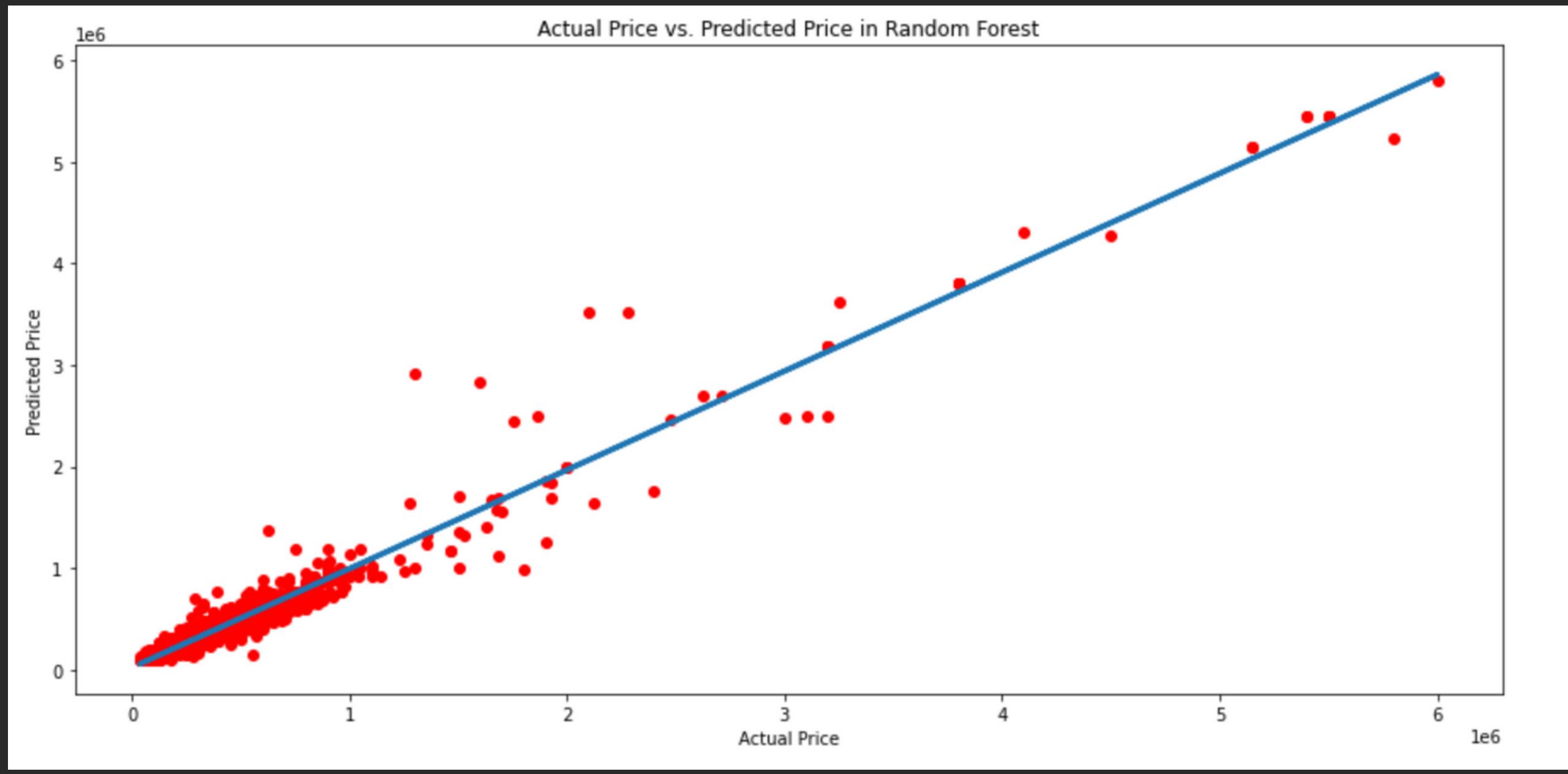
RMSE: 280214.2995

MAPE: 0.4072

R²Score: 0.8742

Random Forest Regression

class = 15, max leaf nodes = 100, max depth = 50



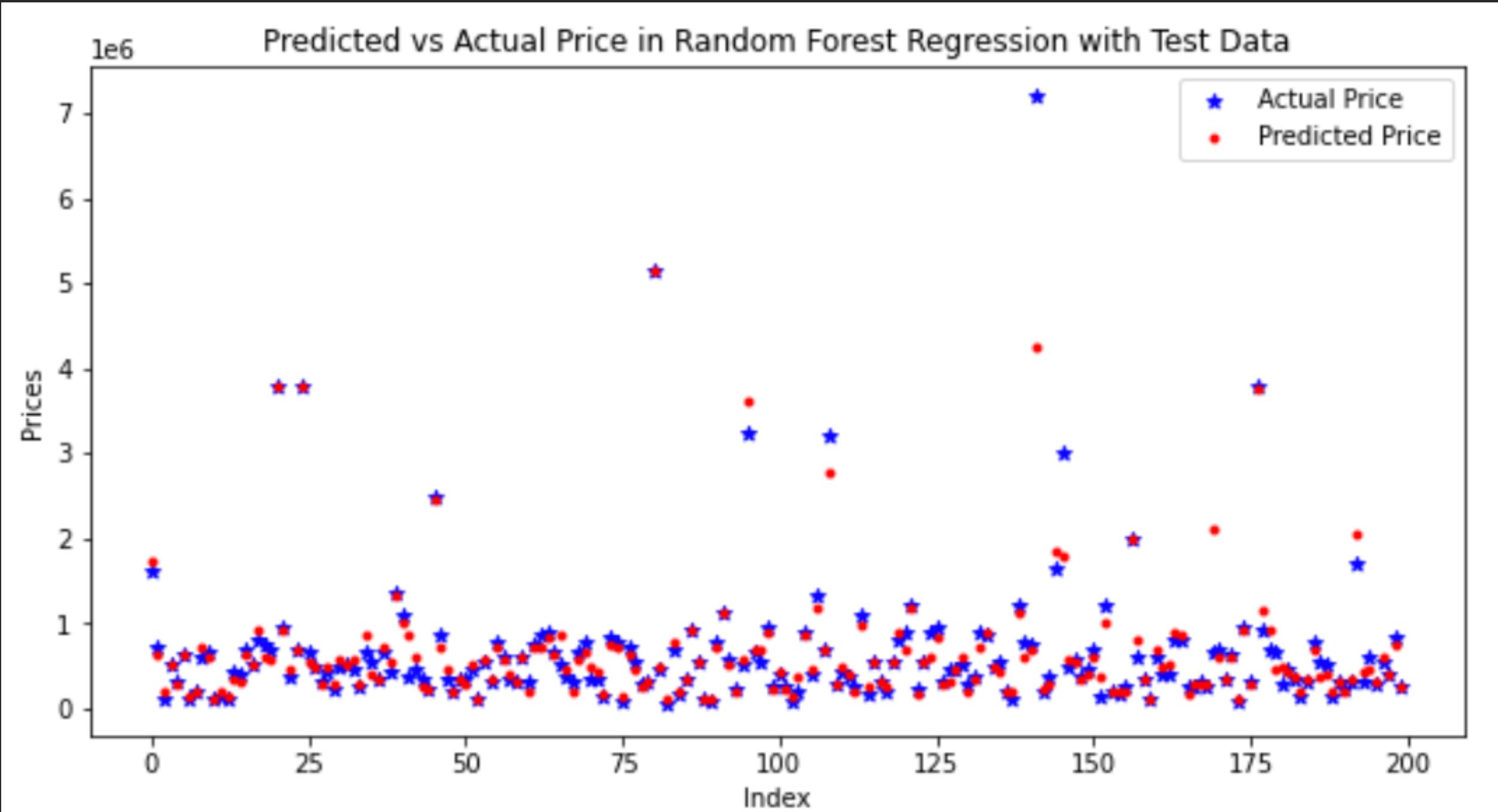
SCORES ON VALIDATION DATA

RMSE: 158744.19254

MAPE: 0.1811

R² Score: 0.9581

Random Forest Regression



SCORES ON TEST DATA

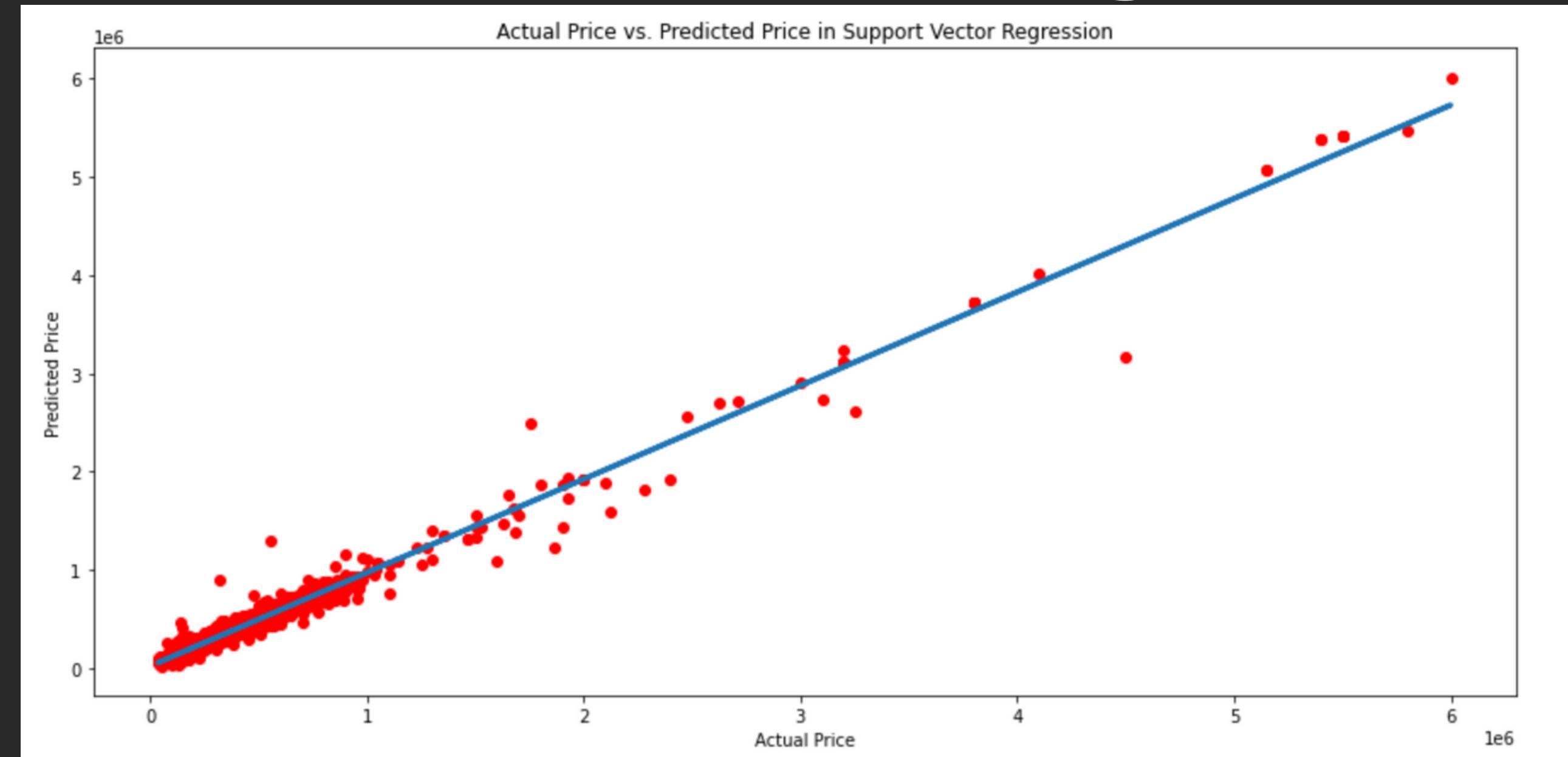
RMSE: 214339.4650

MAPE: 0.1698

R² Score: 0.9146

Support Vector Regression

NO NEGATIVE
PRICE VLAUES



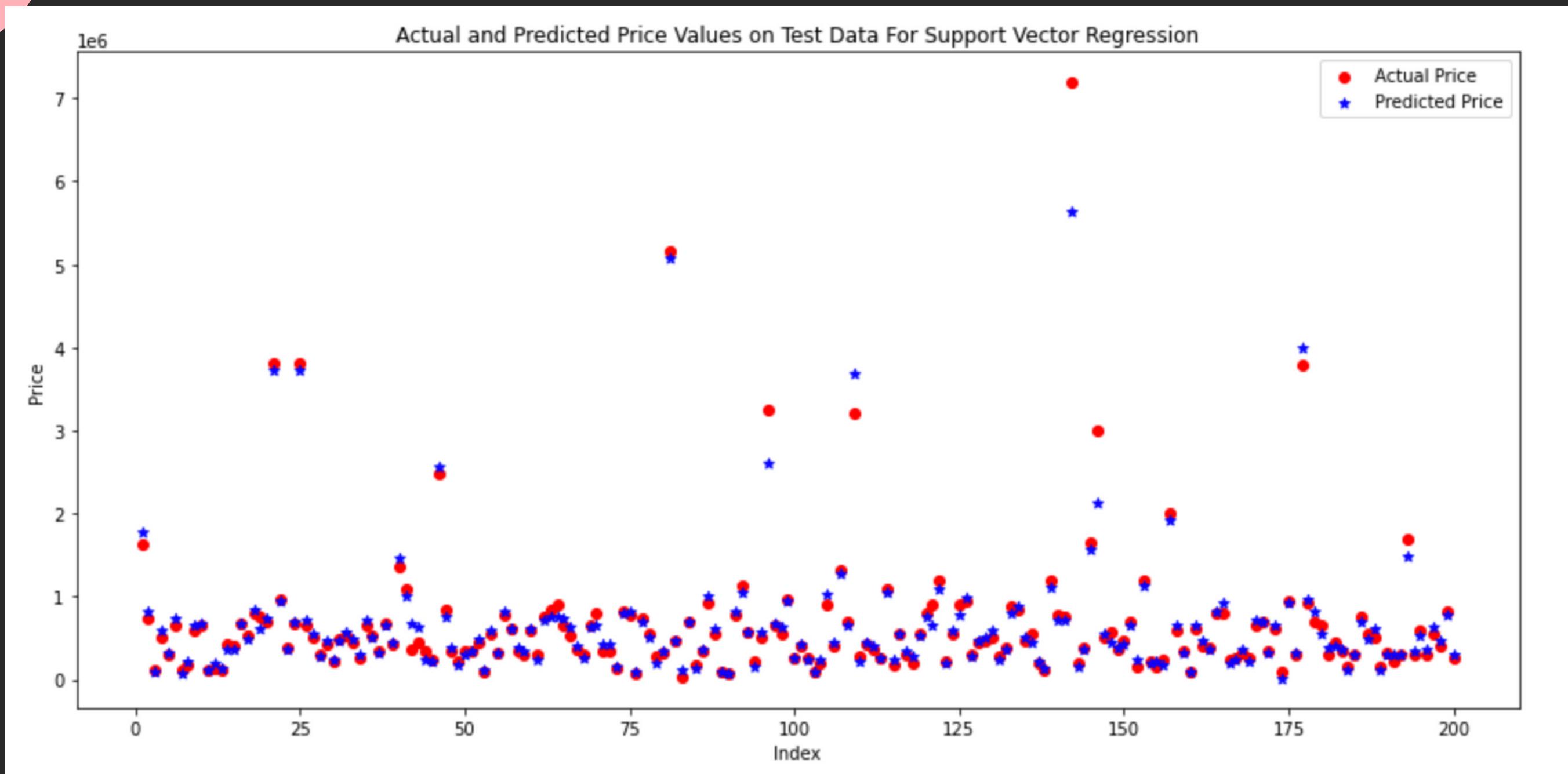
SCORES ON VALIDATION DATA

Mean Squared Error: 107728.4355

MAPE: 0.1580

R^2 Score: 0.9809

Support Vector Regression



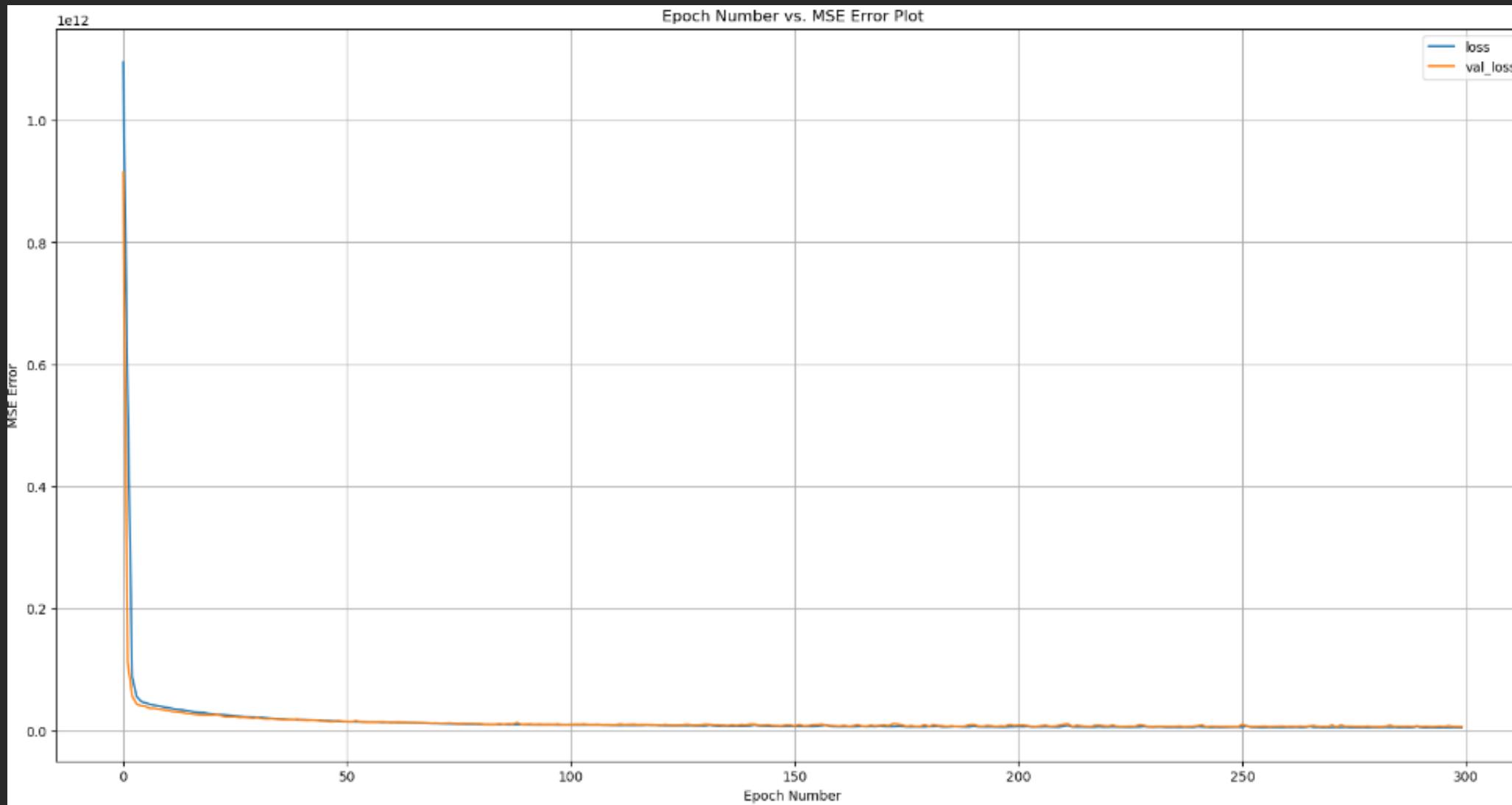
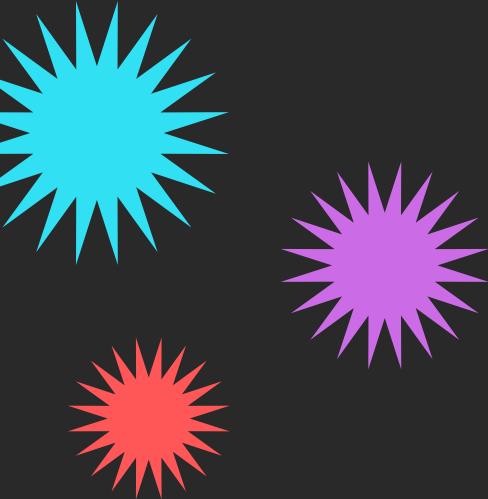
SCORES ON TEST DATA

Mean Squared Error: 215669.4806

MAPE: 0.1499

R²Score: 0.9254

Neural Network



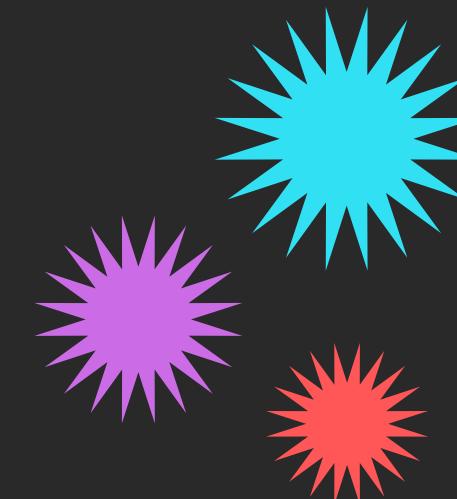
Parameter Adjustments

- hidden layers = [2, 3]
- units = [2, 4, 8, 16, 32, 64, 128]
- epoch = [300]

Initial Model

```
model_dl_small=Sequential()
model_dl_small.add(Dense(32,input_dim=x_train_new.shape[1],activation='relu'))
model_dl_small.add(Dense(16,activation='relu'))
model_dl_small.add(Dense(8,activation='relu'))
model_dl_small.add(Dense(1,activation='linear'))
```

- Adam Optimizer



Neural Network

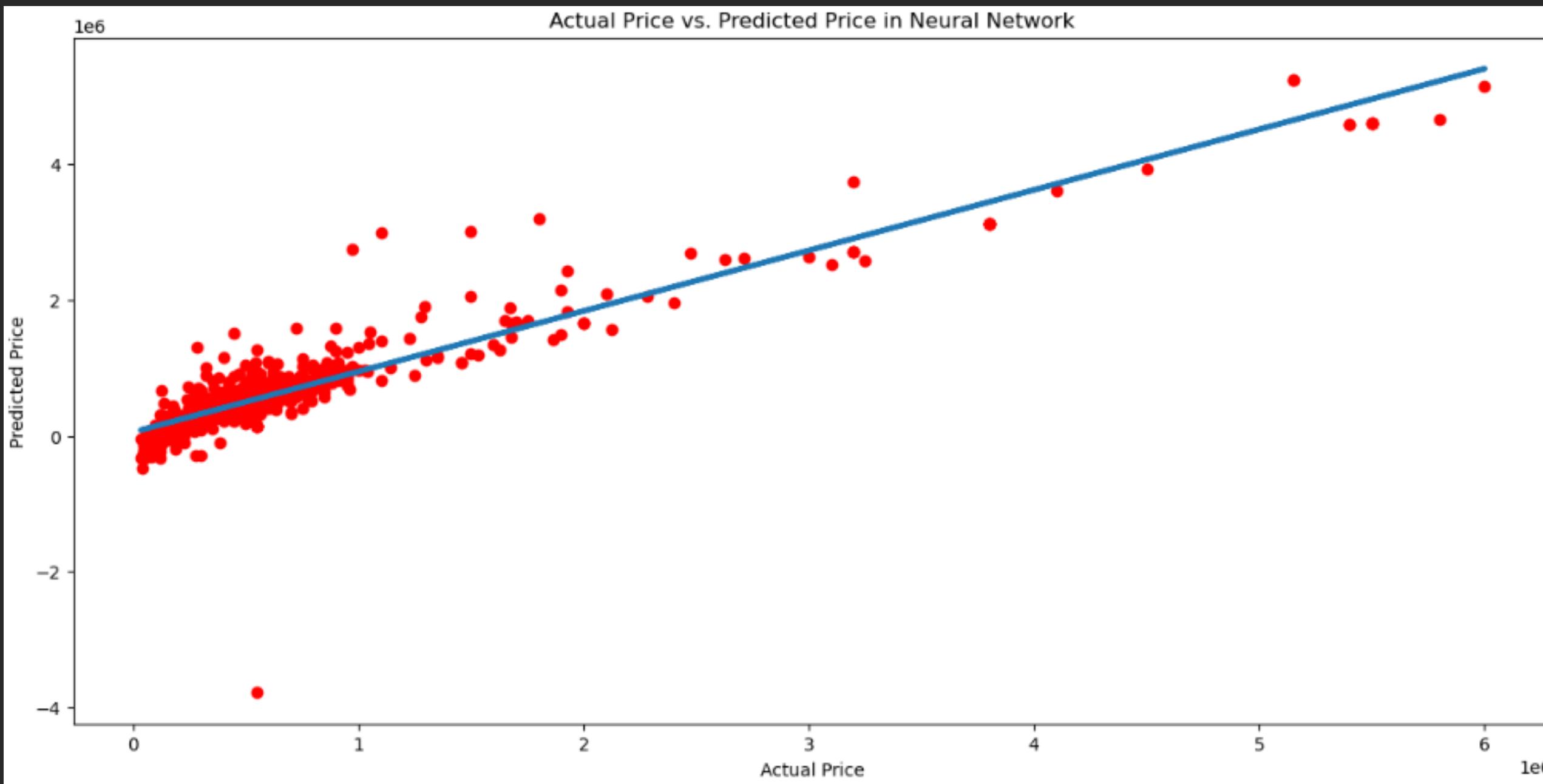
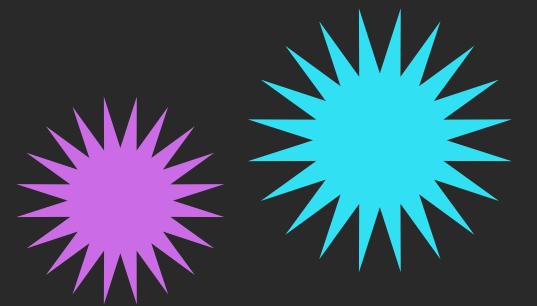
Adjusting Learning Rate on the Initial Model

Learning Rate	0.001	0.005	0.008	0.01
Val R ²	0.9827	0.9935	0.9953	0.9955
Test R ²	0.9689	0.9313	0.9774	0.9570

Adjusting Batch Size of the Model, Learning Rate = 0.008

Batch Size	16	32	64
Val R ²	0.9928	0.9953	0.9876
Test R ²	0.9465	0.9774	0.9482

Neural Network

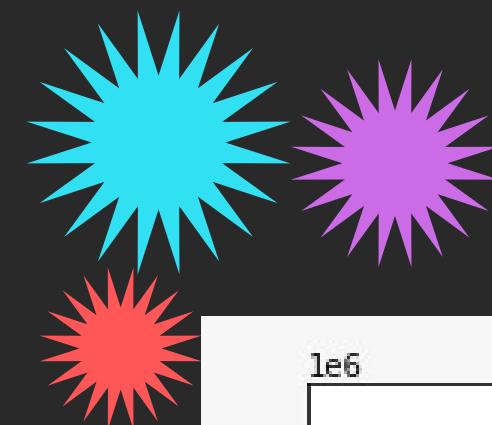


SCORES ON VALIDATION DATA

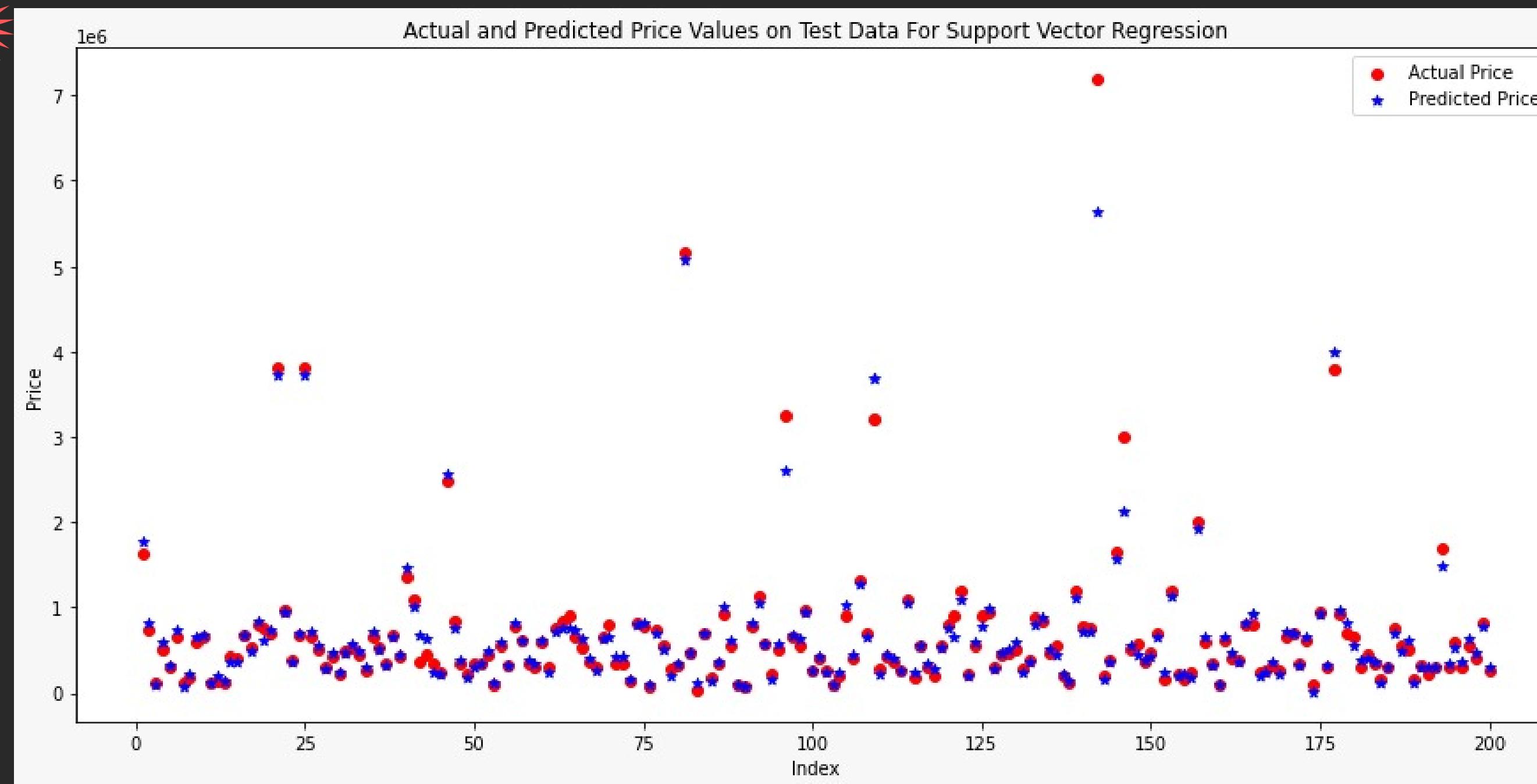
RMSE: 71474.5870

MAPE: 0.1291

R² Score: 0.9953



Neural Network



SCORES ON TEST DATA

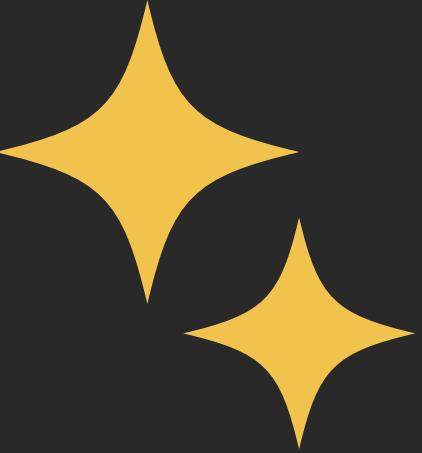
RMSE: 131911.1992

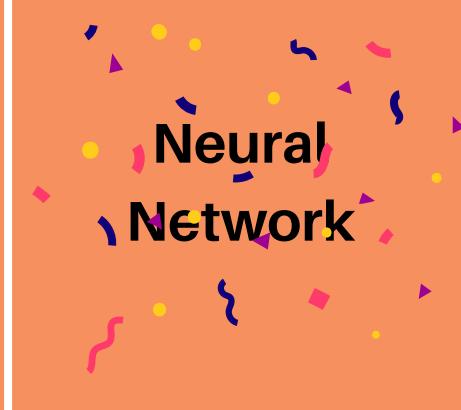
MAPE: 0.0552

R² Score: 0.9774



Comparison on Validation Dataset



	Linear Regression	Ridge Regression	Lasso Regression	Elastic Net	Random Forest		Support Vector Regression
RMSE	296,542.0347	296,533.3638	296,541.2904	293,730.5070	158,744.1925	71,474.5870	107,728.4355
MAPE	0.5023	0.5022	0.5023	0.4777	0.1811	0.1291	0.1580
R^2	0.8553	0.8553	0.8553	0.8580	0.9581	0.9953	0.9809



NEURAL NETWORK GIVES THE BEST RESULT !!

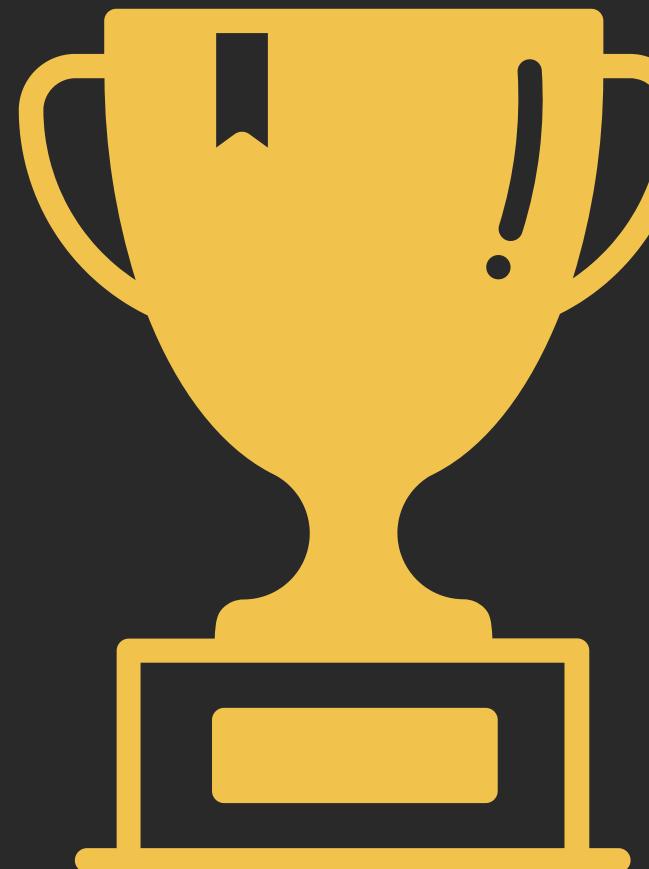


SCORES ON VALIDATION DATA

RMSE: 71474.5870

MAPE: 0.1291

R² Score: 0.9953



SCORES ON TEST DATA

RMSE: 131911.1992

MAPE: 0.0552

R² Score: 0.9774

Thank you!

