Model training

after completed preprocessing and data handling, i have total 12 colums are,

* airline
* flight
* source\_city
* departure\_time
* stops
* arrival\_time
* destination\_city
* class
* duration
* days\_left
* price
* travel\_route

as feture in model i used columns are,

* airline (one hot encoding)
* ~~flight~~
* ~~source\_city~~
* departure\_time (one hot encoding)
* stops
* arrival\_time (one hot encoding)
* ~~destination\_city~~
* class
* duration
* days\_left
* travel\_route (one hot encoding)

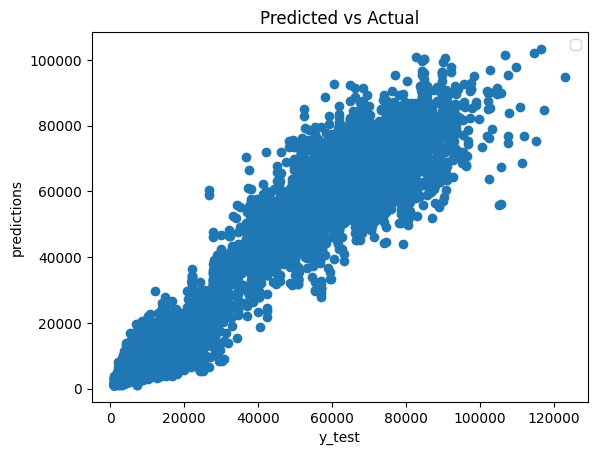
and target is price.

Linear regression Model,

| Metric | Value |
| --- | --- |
| Model score | 0.91063512 |
| MAE | 4531.99925 |
| MSE | 46042761.31 |
| RMSE | 7487.24903 |

Random forest regression model ,

| Metric | Value |
| --- | --- |
| Score on testing data | 0.98524065 |
| Score on training data | 0.99749791 |
| MAE | 1108.28078 |
| MSE | 7659919.02 |
| RMSE | 2767.65587 |



Grid search with parameters are

* 'n\_estimators': [50, 100, 150], # number of trees in the forest
* 'max\_features': [None, 'log2'], # maximum number of features to consider at each split

also use cross validation,

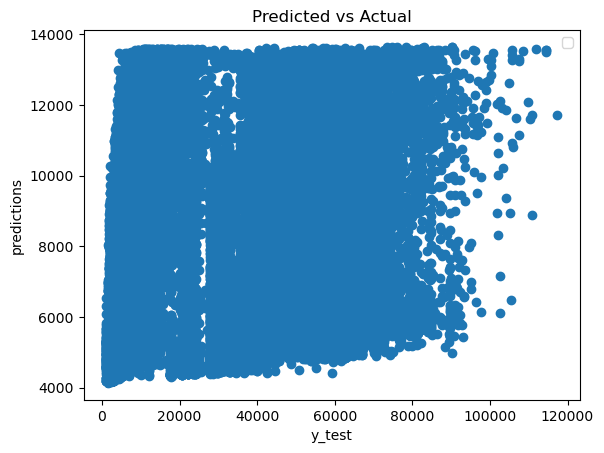
|  |  |
| --- | --- |
| Model | randomforest\_regression |
| Best Parameters | {'max\_features': None, 'n\_estimators': 150} |
| Best Score | 0.984992 |
| Root Mean Squared Error | 1152.122032 |
| Mean Absolute Error | 418.79273 |

Support vector machine regression,

MAE:- 15477.853227362479

MSE:- 674074789.8099297

RMSE:- 25962.95032945851



lasso regression,

score on the testing data :- 0.9112628075952065

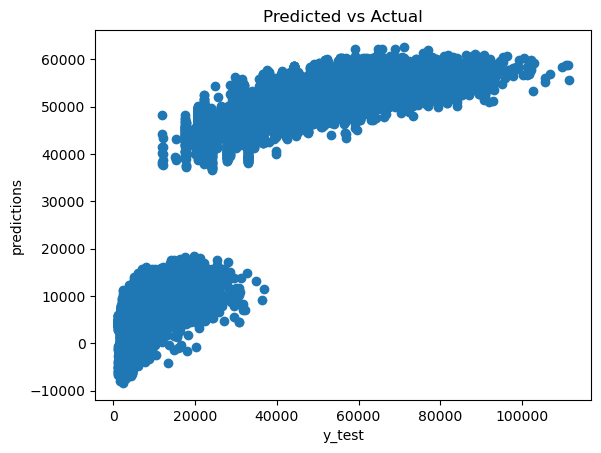
score on the training data :- 0.9115059637386868

MAE:- 4573.318753719743

MSE:- 45684015.912985146

RMSE:- 6758.9951851577125

r2 0.9112628075952065



random forest label encoding :-

x\_train shape:- (210100, 9)

x\_test shape:- (90044, 9)

y\_train shape:- (210100,)

y\_test shape:- (90044,)

model name :- random forest regression

**score on the testing data :- 0.9852479797846426**

**score on the training data :- 0.9975298778733467**

**MAE:- 1102.3006496566325**

**MSE:- 7592783.297002268**

**RMSE:- 2755.5005528945676**

**r2 0.9852479797846426**

Execution time: 37.645846366882324 seconds

random forest one hot encoding:-

x\_train shape:- (210100, 37)

x\_test shape:- (90044, 37)

y\_train shape:- (210100,)

y\_test shape:- (90044,)

model name :- random forest regression

**score on the testing data :- 0.9852483026950065**

**score on the training data :- 0.9975403765074339**

**MAE:- 1087.4013650049224**

**MSE:- 7592617.096822108**

**RMSE:- 2755.470394836807**

**r2 0.9852483026950065**

Execution time: 55.797481536865234 seconds

using flight column,(one hot encoding)

+---------------------------------------+-----------------+

| metrics | value |

+=======================================+=================+

| MAE | 878.238 |

+---------------------------------------+-----------------+

| MSE | 4.98354e+06 |

+---------------------------------------+-----------------+

| RMSE | 2232.38 |

+---------------------------------------+-----------------+

| r2 | 0.990191 |

+---------------------------------------+-----------------+

| explained\_variance\_score | 0.990194 |

+---------------------------------------+-----------------+

| max\_error | 31375.2 |

+---------------------------------------+-----------------+

| mean\_absolute\_percentage\_error | 0.0614783 |

+---------------------------------------+-----------------+

| mean\_squared\_log\_error | 0.0158898 |

+---------------------------------------+-----------------+

| mean\_squared\_log\_error(squared=False) | 0.126055 |

+---------------------------------------+-----------------+

using flight column,(label encoding)

+---------------------------------------+-----------------+

| metrics | value |

+=======================================+=================+

| MAE | 893.228 |

+---------------------------------------+-----------------+

| MSE | 5.11995e+06 |

+---------------------------------------+-----------------+

| RMSE | 2262.73 |

+---------------------------------------+-----------------+

| r2 | 0.989923 |

+---------------------------------------+-----------------+

| explained\_variance\_score | 0.989926 |

+---------------------------------------+-----------------+

| max\_error | 33803 |

+---------------------------------------+-----------------+

| mean\_absolute\_percentage\_error | 0.0622504 |

+---------------------------------------+-----------------+

| mean\_squared\_log\_error | 0.015971 |

+---------------------------------------+-----------------+

| mean\_squared\_log\_error(squared=False) | 0.126376 |

+---------------------------------------+-----------------+

model stacking results,(with flight column)

+---------------------------------------+-----------------+

| metrics | value |

+=======================================+=================+

| MAE | 897.644 |

+---------------------------------------+-----------------+

| MSE | 4.89872e+06 |

+---------------------------------------+-----------------+

| RMSE | 2213.3 |

+---------------------------------------+-----------------+

| r2 | 0.990314 |

+---------------------------------------+-----------------+

| explained\_variance\_score | 0.99032 |

+---------------------------------------+-----------------+

| max\_error | 33426.5 |

+---------------------------------------+-----------------+

| mean\_absolute\_percentage\_error | 0.0623615 |

+---------------------------------------+-----------------+

| mean\_squared\_log\_error | 0.0156323 |

+---------------------------------------+-----------------+

| mean\_squared\_log\_error(squared=False) | 0.125029 |

+---------------------------------------+-----------------+