Flight ticket price prediction

March 28, 2021

1 Predict The Flight Ticket Price

Flight ticket prices can be something hard to guess, today we might see a price, check out the price of the same flight tomorrow, it will be a different story. We might have often heard travellers saying that flight ticket prices are so unpredictable. Huh! Here we take on the challenge! As data scientists, we are gonna prove that given the right data anything can be predicted. Here you will be provided with prices of flight tickets for various airlines between the months of March and June of 2019 and between various cities.

[129]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

[130]:
```
train_df=pd.read_excel('Data_Train.xlsx')
test_df=pd.read_excel('Test_set.xlsx')
```

[131]:
```
train_df.info
```

[131]:
```
<bound method DataFrame.info of            Airline Date_of_Journey    Source
Destination  \
0           IndiGo      24/03/2019  Banglore   New Delhi
1        Air India       1/05/2019   Kolkata    Banglore
2      Jet Airways       9/06/2019     Delhi      Cochin
3           IndiGo      12/05/2019   Kolkata    Banglore
4           IndiGo      01/03/2019  Banglore   New Delhi
…            …             …       …         …
10678     Air Asia       9/04/2019   Kolkata    Banglore
10679    Air India      27/04/2019   Kolkata    Banglore
10680  Jet Airways      27/04/2019  Banglore       Delhi
10681      Vistara      01/03/2019  Banglore   New Delhi
10682    Air India       9/05/2019     Delhi      Cochin

                       Route Dep_Time  Arrival_Time Duration Total_Stops  \
0                  BLR → DEL    22:20  01:10 22 Mar   2h 50m    non-stop
1      CCU → IXR → BBI → BLR    05:50         13:15   7h 25m     2 stops
2      DEL → LKO → BOM → COK    09:25  04:25 10 Jun      19h     2 stops
```

```
3            CCU → NAG → BLR    18:05         23:30   5h 25m      1 stop
4            BLR → NAG → DEL    16:50         21:35   4h 45m      1 stop
…                      …      …           …      …          …
10678              CCU → BLR    19:55         22:25   2h 30m    non-stop
10679              CCU → BLR    20:45         23:20   2h 35m    non-stop
10680              BLR → DEL    08:20         11:20       3h    non-stop
10681              BLR → DEL    11:30         14:10   2h 40m    non-stop
10682  DEL → GOI → BOM → COK    10:55         19:15   8h 20m     2 stops

      Additional_Info  Price
0             No info   3897
1             No info   7662
2             No info  13882
3             No info   6218
4             No info  13302
…               …    …
10678         No info   4107
10679         No info   4145
10680         No info   7229
10681         No info  12648
10682         No info  11753

[10683 rows x 11 columns]>
```

[132]:
```
train_df.columns
```

[132]:
```
Index(['Airline', 'Date_of_Journey', 'Source', 'Destination', 'Route',
       'Dep_Time', 'Arrival_Time', 'Duration', 'Total_Stops',
       'Additional_Info', 'Price'],
      dtype='object')
```

[133]:
```
train_df.shape
```

[133]:
```
(10683, 11)
```

[134]:
```
train_df.head()
```

[134]:
```
       Airline Date_of_Journey    Source Destination                  Route  \
0       IndiGo      24/03/2019  Banglore   New Delhi              BLR → DEL
1    Air India       1/05/2019   Kolkata    Banglore  CCU → IXR → BBI → BLR
2  Jet Airways       9/06/2019     Delhi      Cochin  DEL → LKO → BOM → COK
3       IndiGo      12/05/2019   Kolkata    Banglore        CCU → NAG → BLR
4       IndiGo      01/03/2019  Banglore   New Delhi        BLR → NAG → DEL

  Dep_Time  Arrival_Time Duration Total_Stops Additional_Info  Price
0    22:20  01:10 22 Mar   2h 50m    non-stop         No info   3897
1    05:50         13:15   7h 25m     2 stops         No info   7662
```

```
2    09:25  04:25 10 Jun      19h     2 stops         No info  13882
3    18:05         23:30   5h 25m      1 stop         No info   6218
4    16:50         21:35   4h 45m      1 stop         No info  13302
```

[135]:
```
train_df.isnull().sum()
```

[135]:
```
Airline            0
Date_of_Journey    0
Source             0
Destination        0
Route              1
Dep_Time           0
Arrival_Time       0
Duration           0
Total_Stops        1
Additional_Info    0
Price              0
dtype: int64
```

[136]:
```
train_df.dropna(inplace=True)
```

[137]:
```
train_df.isnull().sum()
```

[137]:
```
Airline            0
Date_of_Journey    0
Source             0
Destination        0
Route              0
Dep_Time           0
Arrival_Time       0
Duration           0
Total_Stops        0
Additional_Info    0
Price              0
dtype: int64
```

[138]:
```
test_df.head()
```

[138]:
```
             Airline Date_of_Journey    Source Destination            Route  \
0        Jet Airways       6/06/2019     Delhi      Cochin  DEL → BOM → COK
1             IndiGo      12/05/2019   Kolkata    Banglore  CCU → MAA → BLR
2        Jet Airways      21/05/2019     Delhi      Cochin  DEL → BOM → COK
3  Multiple carriers      21/05/2019     Delhi      Cochin  DEL → BOM → COK
4           Air Asia      24/06/2019  Banglore       Delhi        BLR → DEL

  Dep_Time  Arrival_Time Duration Total_Stops              Additional_Info
0    17:30  04:25 07 Jun  10h 55m      1 stop                      No info
```

```
1    06:20         10:20       4h       1 stop                    No info
2    19:15  19:00 22 May  23h 45m       1 stop  In-flight meal not included
3    08:00         21:00      13h       1 stop                    No info
4    23:55  02:45 25 Jun   2h 50m     non-stop                    No info
```

[139]:
```
big_df = train_df.append(test_df,sort=False)
```

[140]:
```
big_df.tail()
```

[140]:
```
                Airline Date_of_Journey   Source Destination            Route  \
2666          Air India       6/06/2019  Kolkata    Banglore  CCU → DEL → BLR
2667             IndiGo      27/03/2019  Kolkata    Banglore        CCU → BLR
2668        Jet Airways       6/03/2019    Delhi      Cochin  DEL → BOM → COK
2669          Air India       6/03/2019    Delhi      Cochin  DEL → BOM → COK
2670  Multiple carriers      15/06/2019    Delhi      Cochin  DEL → BOM → COK

     Dep_Time  Arrival_Time Duration Total_Stops Additional_Info  Price
2666    20:30  20:25 07 Jun  23h 55m      1 stop         No info    NaN
2667    14:20         16:55   2h 35m    non-stop         No info    NaN
2668    21:50  04:25 07 Mar   6h 35m      1 stop         No info    NaN
2669    04:00         19:15  15h 15m      1 stop         No info    NaN
2670    04:55         19:15  14h 20m      1 stop         No info    NaN
```

[141]:
```
big_df.dtypes
```

[141]:
```
Airline             object
Date_of_Journey     object
Source              object
Destination         object
Route               object
Dep_Time            object
Arrival_Time        object
Duration            object
Total_Stops         object
Additional_Info     object
Price              float64
dtype: object
```

2 Feature Engineering

[143]:
```
big_df['Date']=big_df['Date_of_Journey'].str.split('/').str[0]
big_df['Month']=big_df['Date_of_Journey'].str.split('/').str[1]
big_df['Year']=big_df['Date_of_Journey'].str.split('/').str[2]
```

[144]:
```
big_df.head()
```

```
[144]:        Airline Date_of_Journey    Source Destination                  Route  \
       0       IndiGo      24/03/2019  Banglore   New Delhi              BLR → DEL
       1    Air India       1/05/2019   Kolkata    Banglore  CCU → IXR → BBI → BLR
       2  Jet Airways       9/06/2019     Delhi      Cochin  DEL → LKO → BOM → COK
       3       IndiGo      12/05/2019   Kolkata    Banglore        CCU → NAG → BLR
       4       IndiGo      01/03/2019  Banglore   New Delhi        BLR → NAG → DEL

         Dep_Time  Arrival_Time Duration Total_Stops Additional_Info    Price Date  \
       0    22:20  01:10 22 Mar   2h 50m    non-stop         No info   3897.0   24
       1    05:50         13:15   7h 25m     2 stops         No info   7662.0    1
       2    09:25  04:25 10 Jun      19h     2 stops         No info  13882.0    9
       3    18:05         23:30   5h 25m      1 stop         No info   6218.0   12
       4    16:50         21:35   4h 45m      1 stop         No info  13302.0   01

         Month  Year
       0    03  2019
       1    05  2019
       2    06  2019
       3    05  2019
       4    03  2019
```

```
[145]: big_df['Date']=big_df['Date'].astype(int)
       big_df['Month']=big_df['Month'].astype(int)
       big_df['Year']=big_df['Year'].astype(int)
```

```
[146]: big_df.dtypes
```

```
[146]: Airline             object
       Date_of_Journey     object
       Source              object
       Destination         object
       Route               object
       Dep_Time            object
       Arrival_Time        object
       Duration            object
       Total_Stops         object
       Additional_Info     object
       Price              float64
       Date                 int32
       Month                int32
       Year                 int32
       dtype: object
```

```
[147]: big_df =big_df.drop(['Date_of_Journey'],axis=1)
```

```
[148]: big_df.head()
```

```
[148]:        Airline    Source Destination                  Route Dep_Time  \
       0       IndiGo  Banglore   New Delhi              BLR → DEL    22:20
       1    Air India   Kolkata    Banglore  CCU → IXR → BBI → BLR    05:50
       2  Jet Airways     Delhi      Cochin  DEL → LKO → BOM → COK    09:25
       3       IndiGo   Kolkata    Banglore        CCU → NAG → BLR    18:05
       4       IndiGo  Banglore   New Delhi        BLR → NAG → DEL    16:50

         Arrival_Time Duration Total_Stops Additional_Info    Price  Date  Month  \
       0 01:10 22 Mar   2h 50m    non-stop         No info   3897.0    24      3
       1        13:15   7h 25m     2 stops         No info   7662.0     1      5
       2 04:25 10 Jun      19h     2 stops         No info  13882.0     9      6
       3        23:30   5h 25m      1 stop         No info   6218.0    12      5
       4        21:35   4h 45m      1 stop         No info  13302.0     1      3

         Year
       0 2019
       1 2019
       2 2019
       3 2019
       4 2019
```

```
[149]: big_df['Arrival_Time']=big_df['Arrival_Time'].str.split(' ').str[0]
```

```
[150]: big_df['Total_Stops']=big_df['Total_Stops'].replace('non-stop','0 stop')
```

```
[151]: big_df.head()
```

```
[151]:        Airline    Source Destination                  Route Dep_Time  \
       0       IndiGo  Banglore   New Delhi              BLR → DEL    22:20
       1    Air India   Kolkata    Banglore  CCU → IXR → BBI → BLR    05:50
       2  Jet Airways     Delhi      Cochin  DEL → LKO → BOM → COK    09:25
       3       IndiGo   Kolkata    Banglore        CCU → NAG → BLR    18:05
       4       IndiGo  Banglore   New Delhi        BLR → NAG → DEL    16:50

        Arrival_Time Duration Total_Stops Additional_Info    Price  Date  Month  \
       0       01:10   2h 50m      0 stop         No info   3897.0    24      3
       1       13:15   7h 25m     2 stops         No info   7662.0     1      5
       2       04:25      19h     2 stops         No info  13882.0     9      6
       3       23:30   5h 25m      1 stop         No info   6218.0    12      5
       4       21:35   4h 45m      1 stop         No info  13302.0     1      3

         Year
       0 2019
       1 2019
       2 2019
       3 2019
       4 2019
```

[152]:

```
big_df['Stop']= big_df['Total_Stops'].str.split(' ').str[0]
```

[153]:

```
big_df.head()
```

[153]:

```
       Airline    Source Destination                  Route Dep_Time  \
0       IndiGo  Banglore   New Delhi              BLR → DEL    22:20
1    Air India   Kolkata    Banglore  CCU → IXR → BBI → BLR    05:50
2  Jet Airways     Delhi      Cochin  DEL → LKO → BOM → COK    09:25
3       IndiGo   Kolkata    Banglore        CCU → NAG → BLR    18:05
4       IndiGo  Banglore   New Delhi        BLR → NAG → DEL    16:50

  Arrival_Time Duration Total_Stops Additional_Info    Price  Date  Month  \
0        01:10   2h 50m      0 stop         No info   3897.0    24      3
1        13:15   7h 25m     2 stops         No info   7662.0     1      5
2        04:25      19h     2 stops         No info  13882.0     9      6
3        23:30   5h 25m      1 stop         No info   6218.0    12      5
4        21:35   4h 45m      1 stop         No info  13302.0     1      3

   Year Stop
0  2019    0
1  2019    2
2  2019    2
3  2019    1
4  2019    1
```

[154]:

```
 big_df.dtypes
```

[154]:

```
Airline             object
Source              object
Destination         object
Route               object
Dep_Time            object
Arrival_Time        object
Duration            object
Total_Stops         object
Additional_Info     object
Price              float64
Date                 int32
Month                int32
Year                 int32
Stop                object
dtype: object
```

[155]:

```
big_df['Stop']=big_df['Stop'].astype(int)
big_df=big_df.drop(['Total_Stops'],axis=1)
```

[156]:

```
big_df.head()
```

[156]:
```
       Airline    Source Destination                  Route Dep_Time  \
0       IndiGo  Banglore   New Delhi              BLR → DEL    22:20
1    Air India   Kolkata    Banglore  CCU → IXR → BBI → BLR    05:50
2  Jet Airways     Delhi      Cochin  DEL → LKO → BOM → COK    09:25
3       IndiGo   Kolkata    Banglore        CCU → NAG → BLR    18:05
4       IndiGo  Banglore   New Delhi        BLR → NAG → DEL    16:50

  Arrival_Time Duration Additional_Info    Price  Date  Month  Year  Stop
0        01:10   2h 50m         No info   3897.0    24      3  2019     0
1        13:15   7h 25m         No info   7662.0     1      5  2019     2
2        04:25      19h         No info  13882.0     9      6  2019     2
3        23:30   5h 25m         No info   6218.0    12      5  2019     1
4        21:35   4h 45m         No info  13302.0     1      3  2019     1
```

[157]:
```python
big_df['Arrival_Hour']=big_df['Arrival_Time'].str.split(':').str[0]
big_df['Arrival_Min']=big_df['Arrival_Time'].str.split(':').str[1]
big_df['Dep_Hour']=big_df['Dep_Time'].str.split(':').str[0]
big_df['Dep_Min']=big_df['Dep_Time'].str.split(':').str[1]
```

[160]:
```python
big_df['Arrival_Hour']=big_df['Arrival_Hour'].astype(int)
big_df['Arrival_Min']=big_df['Arrival_Min'].astype(int)
big_df['Dep_Hour']=big_df['Dep_Hour'].astype(int)
big_df['Dep_Min']=big_df['Dep_Min'].astype(int)
```

[161]:
```python
big_df.dtypes
```

[161]:
```
Airline             object
Source              object
Destination         object
Route               object
Dep_Time            object
Arrival_Time        object
Duration            object
Additional_Info     object
Price              float64
Date                 int32
Month                int32
Year                 int32
Stop                 int32
Arrival_Hour         int32
Arrival_Min          int32
Dep_Hour             int32
Dep_Min              int32
dtype: object
```

[162]:
```python
big_df=big_df.drop(['Arrival_Time'],axis=1)
big_df=big_df.drop(['Dep_Time'],axis=1)
```

[163]:
```
big_df.head()
```

[163]:
```
       Airline    Source Destination                  Route Duration  \
0       IndiGo  Banglore   New Delhi              BLR → DEL   2h 50m
1    Air India   Kolkata    Banglore  CCU → IXR → BBI → BLR   7h 25m
2  Jet Airways     Delhi      Cochin  DEL → LKO → BOM → COK      19h
3       IndiGo   Kolkata    Banglore        CCU → NAG → BLR   5h 25m
4       IndiGo  Banglore   New Delhi        BLR → NAG → DEL   4h 45m

  Additional_Info    Price  Date  Month  Year  Stop  Arrival_Hour  \
0         No info   3897.0    24      3  2019     0             1
1         No info   7662.0     1      5  2019     2            13
2         No info  13882.0     9      6  2019     2             4
3         No info   6218.0    12      5  2019     1            23
4         No info  13302.0     1      3  2019     1            21

   Arrival_Min  Dep_Hour  Dep_Min
0           10        22       20
1           15         5       50
2           25         9       25
3           30        18        5
4           35        16       50
```

[164]:
```
big_df['Route_1']=big_df['Route'].str.split('→ ').str[0]
big_df['Route_2']=big_df['Route'].str.split('→ ').str[1]
big_df['Route_3']=big_df['Route'].str.split('→ ').str[2]
big_df['Route_4']=big_df['Route'].str.split('→ ').str[3]
big_df['Route_5']=big_df['Route'].str.split('→ ').str[4]
```

[165]:
```
big_df.isnull().sum()
```

[165]:
```
Airline               0
Source                0
Destination           0
Route                 0
Duration              0
Additional_Info       0
Price              2671
Date                  0
Month                 0
Year                  0
Stop                  0
Arrival_Hour          0
Arrival_Min           0
Dep_Hour              0
Dep_Min               0
Route_1               0
```

```
Route_2                0
Route_3             4340
Route_4            11396
Route_5            13295
dtype: int64
```

```
[166]: big_df['Price'].fillna((big_df['Price'].mean()),inplace=True)
```

```
[168]: big_df['Route_1'].fillna('None',inplace=True)
       big_df['Route_2'].fillna('None',inplace=True)
       big_df['Route_3'].fillna('None',inplace=True)
       big_df['Route_4'].fillna('None',inplace=True)
       big_df['Route_5'].fillna('None',inplace=True)
```

```
[169]: big_df.head()
```

```
[169]:        Airline    Source Destination                  Route Duration  \
       0       IndiGo  Banglore   New Delhi              BLR → DEL   2h 50m
       1    Air India   Kolkata    Banglore  CCU → IXR → BBI → BLR   7h 25m
       2  Jet Airways     Delhi      Cochin  DEL → LKO → BOM → COK      19h
       3       IndiGo   Kolkata    Banglore        CCU → NAG → BLR   5h 25m
       4       IndiGo  Banglore   New Delhi        BLR → NAG → DEL   4h 45m

         Additional_Info    Price  Date  Month  Year  Stop  Arrival_Hour  \
       0         No info   3897.0    24      3  2019     0             1
       1         No info   7662.0     1      5  2019     2            13
       2         No info  13882.0     9      6  2019     2             4
       3         No info   6218.0    12      5  2019     1            23
       4         No info  13302.0     1      3  2019     1            21

          Arrival_Min  Dep_Hour  Dep_Min Route_1 Route_2 Route_3 Route_4 Route_5
       0           10        22       20    BLR      DEL    None    None    None
       1           15         5       50    CCU     IXR     BBI      BLR    None
       2           25         9       25    DEL     LKO     BOM      COK    None
       3           30        18        5    CCU     NAG      BLR    None    None
       4           35        16       50    BLR     NAG      DEL    None    None
```

```
[171]: big_df=big_df.drop(['Route'],axis=1)
       big_df=big_df.drop(['Duration'],axis=1)
```

```
[172]: big_df.head()
```

```
[172]:        Airline    Source Destination Additional_Info    Price  Date  Month  \
       0       IndiGo  Banglore   New Delhi         No info   3897.0    24      3
       1    Air India   Kolkata    Banglore         No info   7662.0     1      5
       2  Jet Airways     Delhi      Cochin         No info  13882.0     9      6
       3       IndiGo   Kolkata    Banglore         No info   6218.0    12      5
```

```
4       IndiGo  Banglore   New Delhi         No info  13302.0     1      3

   Year  Stop  Arrival_Hour  Arrival_Min  Dep_Hour  Dep_Min Route_1 Route_2   \
0  2019     0             1           10        22       20     BLR     DEL
1  2019     2            13           15         5       50     CCU     IXR
2  2019     2             4           25         9       25     DEL     LKO
3  2019     1            23           30        18        5     CCU     NAG
4  2019     1            21           35        16       50     BLR     NAG

  Route_3 Route_4 Route_5
0    None    None    None
1    BBI      BLR    None
2    BOM      COK    None
3     BLR    None    None
4     DEL    None    None
```

[173]:
```
big_df.isnull().sum()
```

[173]:
```
Airline            0
Source             0
Destination        0
Additional_Info    0
Price              0
Date               0
Month              0
Year               0
Stop               0
Arrival_Hour       0
Arrival_Min        0
Dep_Hour           0
Dep_Min            0
Route_1            0
Route_2            0
Route_3            0
Route_4            0
Route_5            0
dtype: int64
```

[176]:
```
from sklearn.preprocessing import LabelEncoder
encoder=LabelEncoder()
big_df['Airline']=encoder.fit_transform(big_df['Airline'])
big_df["Source"]=encoder.fit_transform(big_df['Source'])
big_df["Destination"]=encoder.fit_transform(big_df['Destination'])
big_df["Additional_Info"]=encoder.fit_transform(big_df['Additional_Info'])
big_df["Route_1"]=encoder.fit_transform(big_df['Route_1'])
big_df["Route_2"]=encoder.fit_transform(big_df['Route_2'])
big_df["Route_3"]=encoder.fit_transform(big_df['Route_3'])
```

```
big_df["Route_4"]=encoder.fit_transform(big_df['Route_4'])
big_df["Route_5"]=encoder.fit_transform(big_df['Route_5'])
```

[177]:
```
big_df.head()
```

[177]:
```
   Airline  Source  Destination  Additional_Info    Price  Date  Month  Year  \
0        3       0            5                8   3897.0    24      3  2019
1        1       3            0                8   7662.0     1      5  2019
2        4       2            1                8  13882.0     9      6  2019
3        3       3            0                8   6218.0    12      5  2019
4        3       0            5                8  13302.0     1      3  2019

   Stop  Arrival_Hour  Arrival_Min  Dep_Hour  Dep_Min  Route_1  Route_2  \
0     0             1           10        22       20        0       13
1     2            13           15         5       50        2       25
2     2             4           25         9       25        3       32
3     1            23           30        18        5        2       34
4     1            21           35        16       50        0       34

   Route_3  Route_4  Route_5
0       24       12        4
1        1        3        4
2        4        5        4
3        3       12        4
4        8       12        4
```

2.0.1 Feature Selection

[179]:
```
from sklearn.linear_model import Lasso
from sklearn.feature_selection import SelectFromModel
```

[185]:
```
big_df.shape
```

[185]:
```
(13353, 18)
```

[182]:
```
df_train=big_df[0:10683]

df_test=big_df[10683:]
```

[183]:
```
df_test
```

[183]:
```
      Airline  Source  Destination  Additional_Info        Price  Date  Month  \
1           3       3            0                8  9087.214567    12      5
2           4       2            1                5  9087.214567    21      5
3           6       2            1                8  9087.214567    21      5
4           0       0            2                8  9087.214567    24      6
5           4       2            1                5  9087.214567    12      6
…       …      …          …              …          …    …     …
```

```
2666        1       3            0                8  9087.214567     6      6
2667        3       3            0                8  9087.214567    27      3
2668        4       2            1                8  9087.214567     6      3
2669        1       2            1                8  9087.214567     6      3
2670        6       2            1                8  9087.214567    15      6

      Year  Stop  Arrival_Hour  Arrival_Min  Dep_Hour  Dep_Min  Route_1  \
1     2019     1            10           20         6       20        2
2     2019     1            19            0        19       15        3
3     2019     1            21            0         8        0        3
4     2019     0             2           45        23       55        0
5     2019     1            12           35        18       15        3
…    …   …           …          …       …      …      …
2666  2019     1            20           25        20       30        2
2667  2019     0            16           55        14       20        2
2668  2019     1             4           25        21       50        3
2669  2019     1            19           15         4        0        3
2670  2019     1            19           15         4       55        3

      Route_2  Route_3  Route_4  Route_5
1          33        3       12        4
2           7        6       12        4
3           7        6       12        4
4          13       24       12        4
5           7        6       12        4
…       …      …      …      …
2666       14        3       12        4
2667        5       24       12        4
2668        7        6       12        4
2669        7        6       12        4
2670        7        6       12        4

[2670 rows x 18 columns]
```

[187]:

```
X=df_train.drop(['Price'],axis=1)
y=df_train.Price
```

[189]:

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test= train_test_split(X,y,test_size=0.
 ↪3,random_state=0)
```

[191]:

```
model=SelectFromModel(Lasso(alpha=0.005,random_state=0))
```

[192]:

```
model.fit(X_train,y_train)
```

[192]:

```
SelectFromModel(estimator=Lasso(alpha=0.005, copy_X=True, fit_intercept=True,
                                max_iter=1000, normalize=False, positive=False,
                                precompute=False, random_state=0,
                                selection='cyclic', tol=0.0001,
                                warm_start=False),
               max_features=None, norm_order=1, prefit=False, threshold=None)
```

```
[194]: model.get_support()
```

```
[194]: array([ True,  True,  True,  True,  True,  True, False,  True,  True,
               True,  True,  True,  True,  True,  True,  True,  True])
```

```
[195]: Selected_features=X_train.columns[(model.get_support())]
```

```
[196]: Selected_features
```

```
[196]: Index(['Airline', 'Source', 'Destination', 'Additional_Info', 'Date', 'Month',
              'Stop', 'Arrival_Hour', 'Arrival_Min', 'Dep_Hour', 'Dep_Min', 'Route_1',
              'Route_2', 'Route_3', 'Route_4', 'Route_5'],
             dtype='object')
```

```
[197]: X_train.drop(['Year'],axis=1)
       X_test.drop(['Year'],axis=1)
```

```
[197]:        Airline  Source  Destination  Additional_Info  Date  Month  Stop  \
       9694         8       0            2                8    15      6     0
       9826         2       0            5                8     3      3     0
       7702         1       3            0                8     6      6     2
       1437         4       0            5                8     6      3     1
       6828         3       2            1                8    15      6     1
       …          …     …           …              …     …     …    …
       2294         3       2            1                8    21      5     1
       7085         3       3            0                8    18      3     0
       10332        4       3            0                5    24      3     1
       872          4       3            0                8    18      5     1
       6935         4       4            3                8    27      6     0

              Arrival_Hour  Arrival_Min  Dep_Hour  Dep_Min  Route_1  Route_2  \
       9694              8           35         5       55        0       13
       9826             23           50        20       55        0       13
       7702             20           25         5       50        2       25
       1437             14           25         9       45        0       33
       6828              1           30        16        0        3        7
       …             …          …       …        …       …       …
       2294             21            0         8       30        3        7
       7085             23            5        20       25        2        5
       10332             4           45        19       45        2        7
       872              10            5        21       10        2        7
       6935             10           15         8       45        1       19
              Route_3  Route_4  Route_5
       9694        24       12        4
       9826        24       12        4
       7702         9        3        4
       1437         8       12        4
       6828         6       12        4
       …         …      …      …
       2294         6       12        4
       7085        24       12        4
       10332        3       12        4
       872          3       12        4
       6935        24       12        4

       [3205 rows x 16 columns]
```

2.0.2 RandomForestRegressor

[208]:

```
from sklearn.model_selection import RandomizedSearchCV

#number of trees in Random Forest
n_estimators=[int(x) for x in np.linspace(start=100,stop=1200,num=12)]

#number of features to consider at every split
max_features=['auto','sqrt']

#maximum num of leaves in tree
max_depth=[int(x) for x in np.linspace(5,30,num=6)]

#min number of samples required to split a node
min_samples_split=[2,5,10,15,100]

#min number of samples required at ech leaf node
min_samples_leaf=[1,2,5,10]
```

[209]:

```
random_grid={'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf}
```

[211]:

```
print(random_grid)
```

```
{'n_estimators': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100,
1200], 'max_features': ['auto', 'sqrt'], 'max_depth': [5, 10, 15, 20, 25, 30],
'min_samples_split': [2, 5, 10, 15, 100], 'min_samples_leaf': [1, 2, 5, 10]}
```

[212]:
```
from sklearn.ensemble import RandomForestRegressor
rf=RandomForestRegressor()
```

[213]:
```
rf_random = RandomizedSearchCV(estimator = rf, param_distributions =␣
 ↪random_grid,scoring='neg_mean_squared_error', n_iter = 50,cv = 5, verbose=2,␣
 ↪random_state=42, n_jobs = 1)
```

[215]:
```
rf_random.fit(X_train,y_train)
```

```
Fitting 5 folds for each of 50 candidates, totalling 250 fits
[CV] n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5

[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.

[CV]  n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5, total=   2.5s
[CV] n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5

[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    2.5s remaining:    0.0s

[CV]  n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5, total=   2.2s
[CV] n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5
[CV]  n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5, total=   2.2s
[CV] n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5
[CV]  n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5, total=   2.3s
[CV] n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5
[CV]  n_estimators=400, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=5, total=   2.5s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20, total=   6.1s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20, total=   6.0s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20, total=   5.4s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20
```

```
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20, total=   5.8s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=20, total=   6.0s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   1.7s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   1.7s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   1.7s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   1.6s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   1.6s
[CV] n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  18.6s
[CV] n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  13.3s
[CV] n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  13.2s
[CV] n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  14.2s
[CV] n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=600, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  15.4s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15
```

```
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   9.1s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   7.4s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   7.3s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   6.9s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   6.9s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   5.7s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   6.0s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   5.8s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   5.9s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   6.2s
[CV] n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15
[CV]  n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   1.6s
[CV] n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15
[CV]  n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   1.7s
[CV] n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15
```

```
[CV]  n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   1.9s
[CV] n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15
[CV]  n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   1.8s
[CV] n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15
[CV]  n_estimators=100, min_samples_split=100, min_samples_leaf=5,
max_features=auto, max_depth=15, total=   1.7s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   9.3s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   9.2s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   9.1s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   9.1s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   9.2s
[CV] n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10
[CV]  n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   8.5s
[CV] n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10
[CV]  n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   7.9s
[CV] n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10
[CV]  n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   7.9s
[CV] n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10
[CV]  n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   7.7s
[CV] n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10
```

```
[CV]  n_estimators=1000, min_samples_split=15, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   7.9s
[CV] n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   1.0s
[CV] n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   0.9s
[CV] n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   1.0s
[CV] n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   1.1s
[CV] n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=100, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   1.0s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=   3.0s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=   3.0s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=   2.9s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=   3.0s
[CV] n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=300, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=   2.8s
[CV] n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30
[CV]  n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30, total=   4.3s
[CV] n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30
```

```
[CV]  n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30, total=   4.0s
[CV] n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30
[CV]  n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30, total=   4.5s
[CV] n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30
[CV]  n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30, total=   4.2s
[CV] n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30
[CV]  n_estimators=400, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=30, total=   4.1s
[CV] n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5
[CV]  n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5, total=   6.2s
[CV] n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5
[CV]  n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5, total=   5.2s
[CV] n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5
[CV]  n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5, total=   6.1s
[CV] n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5
[CV]  n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5, total=   6.0s
[CV] n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5
[CV]  n_estimators=900, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=5, total=   5.4s
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=  10.2s
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=  10.5s
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=  11.1s
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
```

```
[CV]  n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=  11.4s
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=900, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=  10.6s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.6s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.7s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.6s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.6s
[CV] n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.6s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.8s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.8s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.9s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.7s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   1.7s
[CV] n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10
```

```
[CV]  n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10, total=  14.3s
[CV] n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10
[CV]  n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10, total=  14.7s
[CV] n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10
[CV]  n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10, total=  15.2s
[CV] n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10
[CV]  n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10, total=  15.3s
[CV] n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10
[CV]  n_estimators=700, min_samples_split=5, min_samples_leaf=1,
max_features=auto, max_depth=10, total=  14.6s
[CV] n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5
[CV]  n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5, total=  14.8s
[CV] n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5
[CV]  n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5, total=  15.3s
[CV] n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5
[CV]  n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5, total=  15.1s
[CV] n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5
[CV]  n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5, total=  14.4s
[CV] n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5
[CV]  n_estimators=1200, min_samples_split=100, min_samples_leaf=10,
max_features=auto, max_depth=5, total=  15.6s
[CV] n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5
[CV]  n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5, total=   4.8s
[CV] n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5
[CV]  n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5, total=   4.9s
[CV] n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5
```

```
[CV]  n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5, total=   5.1s
[CV] n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5
[CV]  n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5, total=   4.7s
[CV] n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5
[CV]  n_estimators=800, min_samples_split=10, min_samples_leaf=2,
max_features=sqrt, max_depth=5, total=   4.4s
[CV] n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   7.7s
[CV] n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   8.2s
[CV] n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   7.6s
[CV] n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   8.6s
[CV] n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10
[CV]  n_estimators=1100, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=10, total=   8.2s
[CV] n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10
[CV]  n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   3.7s
[CV] n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10
[CV]  n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   4.2s
[CV] n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10
[CV]  n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   4.0s
[CV] n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10
[CV]  n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   3.8s
[CV] n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10
```

```
[CV]  n_estimators=500, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=10, total=   4.2s
[CV] n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=  12.5s
[CV] n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=  11.2s
[CV] n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=  11.2s
[CV] n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=  11.6s
[CV] n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=  14.4s
[CV] n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15, total=  18.0s
[CV] n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15, total=  12.5s
[CV] n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15, total=  11.5s
[CV] n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15, total=  11.3s
[CV] n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15
[CV]  n_estimators=1000, min_samples_split=2, min_samples_leaf=2,
max_features=sqrt, max_depth=15, total=  11.6s
[CV] n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=  10.4s
[CV] n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
```

```
[CV]  n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   9.7s
[CV] n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   9.6s
[CV] n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   9.9s
[CV] n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=1200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=  10.1s
[CV] n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   2.8s
[CV] n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   2.8s
[CV] n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   3.1s
[CV] n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   3.1s
[CV] n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=300, min_samples_split=15, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   2.9s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20, total=  16.0s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20, total=  16.4s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20, total=  16.3s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20
```

```
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20, total=  16.2s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=20, total=  17.1s
[CV] n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25, total=  29.2s
[CV] n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25, total=  30.0s
[CV] n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25, total=  28.7s
[CV] n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25, total=  29.7s
[CV] n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=1100, min_samples_split=5, min_samples_leaf=2,
max_features=auto, max_depth=25, total=  30.2s
[CV] n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15
[CV]  n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15, total=   4.9s
[CV] n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15
[CV]  n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15, total=   5.0s
[CV] n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15
[CV]  n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15, total=   5.0s
[CV] n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15
[CV]  n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15, total=   5.2s
[CV] n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15
[CV]  n_estimators=300, min_samples_split=100, min_samples_leaf=1,
max_features=auto, max_depth=15, total=   5.7s
[CV] n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
```

```
[CV]  n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   1.3s
[CV] n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   1.2s
[CV] n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   1.3s
[CV] n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   1.2s
[CV] n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20
[CV]  n_estimators=100, min_samples_split=5, min_samples_leaf=2,
max_features=sqrt, max_depth=20, total=   1.2s
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20
[CV]  n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20, total=   5.2s
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20
[CV]  n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20, total=   5.0s
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20
[CV]  n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20, total=   5.1s
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20
[CV]  n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20, total=   4.8s
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20
[CV]  n_estimators=700, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=20, total=   4.7s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.4s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.4s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15
```

```
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.4s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.5s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.4s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  11.7s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  11.3s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  11.9s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  11.8s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=20, total=  11.5s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   7.3s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   7.6s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   7.5s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   7.8s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
```

```
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   7.6s
[CV] n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30
[CV]  n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30, total=  26.5s
[CV] n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30
[CV]  n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30, total=  26.6s
[CV] n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30
[CV]  n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30, total=  30.0s
[CV] n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30
[CV]  n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30, total=  26.8s
[CV] n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30
[CV]  n_estimators=1200, min_samples_split=15, min_samples_leaf=5,
max_features=auto, max_depth=30, total=  27.4s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25, total=  22.7s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25, total=  21.9s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25, total=  22.3s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25, total=  22.3s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=25, total=  22.5s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10, total=   5.0s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10
```

```
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10, total=   5.0s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10, total=   5.2s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10, total=   5.0s
[CV] n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10
[CV]  n_estimators=600, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=10, total=   5.4s
[CV] n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25, total=   7.5s
[CV] n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25, total=   8.3s
[CV] n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25, total=   7.6s
[CV] n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25, total=   7.4s
[CV] n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=5, min_samples_leaf=5,
max_features=sqrt, max_depth=25, total=   7.4s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5, total=   5.9s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5, total=   6.1s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5, total=   6.6s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5
```

```
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5, total=   6.4s
[CV] n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5
[CV]  n_estimators=500, min_samples_split=2, min_samples_leaf=5,
max_features=auto, max_depth=5, total=   6.1s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   5.5s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   5.9s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   5.8s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   5.7s
[CV] n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25
[CV]  n_estimators=800, min_samples_split=100, min_samples_leaf=2,
max_features=sqrt, max_depth=25, total=   5.8s
[CV] n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=  11.1s
[CV] n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=  12.8s
[CV] n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=  13.2s
[CV] n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=  11.8s
[CV] n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30
[CV]  n_estimators=1200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=30, total=  11.3s
[CV] n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30
```

```
[CV]  n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.0s
[CV] n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30
[CV]  n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.2s
[CV] n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30
[CV]  n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.4s
[CV] n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30
[CV]  n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.5s
[CV] n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30
[CV]  n_estimators=600, min_samples_split=10, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.7s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20, total=  22.6s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20, total=  23.2s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20, total=  22.5s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20, total=  22.7s
[CV] n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20
[CV]  n_estimators=900, min_samples_split=10, min_samples_leaf=1,
max_features=auto, max_depth=20, total=  22.2s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15, total=   1.9s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15, total=   1.8s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15
```

```
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15, total=   1.8s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15, total=   1.8s
[CV] n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=15, total=   2.1s
[CV] n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   6.9s
[CV] n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   6.2s
[CV] n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   6.5s
[CV] n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   6.0s
[CV] n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25
[CV]  n_estimators=700, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=25, total=   6.1s
[CV] n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.8s
[CV] n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.8s
[CV] n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.8s
[CV] n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.8s
[CV] n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15
```

```
[CV]  n_estimators=200, min_samples_split=10, min_samples_leaf=10,
max_features=sqrt, max_depth=15, total=   1.9s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25, total=   3.6s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25, total=   3.4s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25, total=   3.4s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25, total=   3.6s
[CV] n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25
[CV]  n_estimators=200, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=25, total=   3.4s
[CV] n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   3.7s
[CV] n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   4.1s
[CV] n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   4.0s
[CV] n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   4.2s
[CV] n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20
[CV]  n_estimators=400, min_samples_split=2, min_samples_leaf=5,
max_features=sqrt, max_depth=20, total=   4.3s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5, total=   5.8s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5
```

```
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5, total=   5.3s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5, total=   5.6s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5, total=   5.9s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=5, total=   5.3s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.6s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.5s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.4s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.9s
[CV] n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30
[CV]  n_estimators=900, min_samples_split=100, min_samples_leaf=1,
max_features=sqrt, max_depth=30, total=   6.9s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   2.2s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   2.8s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   2.7s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
```

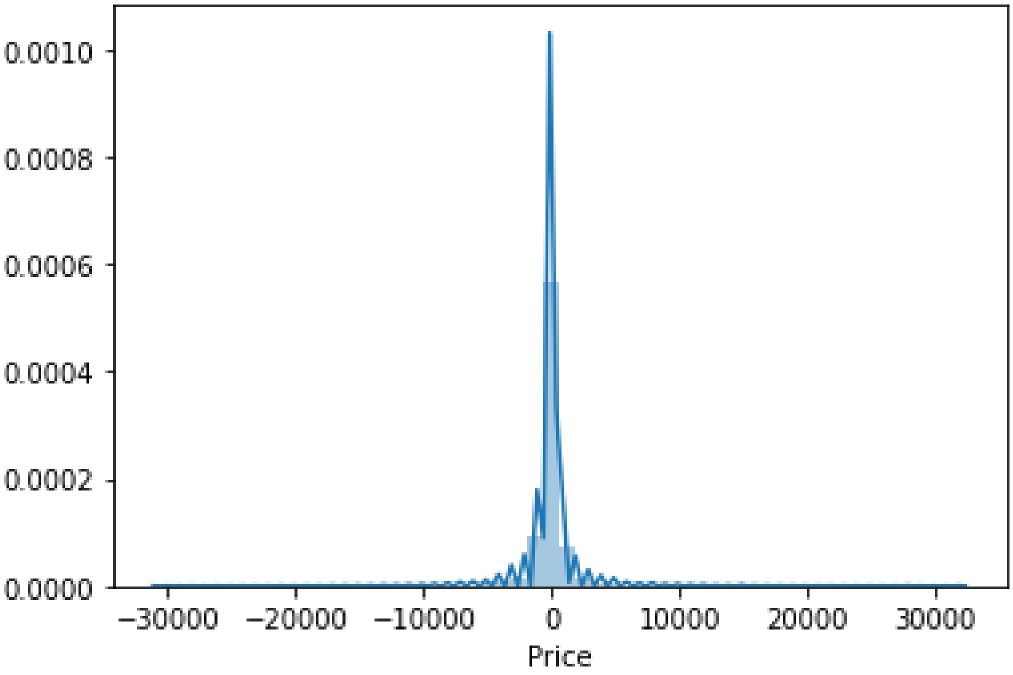
```
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   2.3s
[CV] n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15
[CV]  n_estimators=200, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=15, total=   2.2s

[Parallel(n_jobs=1)]: Done 250 out of 250 | elapsed: 33.0min finished
```

```
[215]: RandomizedSearchCV(cv=5, error_score=nan,
                          estimator=RandomForestRegressor(bootstrap=True,
                                                          ccp_alpha=0.0,
                                                          criterion='mse',
                                                          max_depth=None,
                                                          max_features='auto',
                                                          max_leaf_nodes=None,
                                                          max_samples=None,
                                                          min_impurity_decrease=0.0,
                                                          min_impurity_split=None,
                                                          min_samples_leaf=1,
                                                          min_samples_split=2,
                                                          min_weight_fraction_leaf=0.0,
                                                          n_estimators=100,
                                                          n_jobs=None,
        oob_score=Fals…
                          iid='deprecated', n_iter=50, n_jobs=1,
                          param_distributions={'max_depth': [5, 10, 15, 20, 25, 30],
                                               'max_features': ['auto', 'sqrt'],
                                               'min_samples_leaf': [1, 2, 5, 10],
                                               'min_samples_split': [2, 5, 10, 15,
                                                                     100],
                                               'n_estimators': [100, 200, 300, 400,
                                                                500, 600, 700, 800,
                                                                900, 1000, 1100,
                                                                1200]},
                          pre_dispatch='2*n_jobs', random_state=42, refit=True,
                          return_train_score=False, scoring='neg_mean_squared_error',
                          verbose=2)
```

```
[216]: y_pred=rf_random.predict(X_test)
```

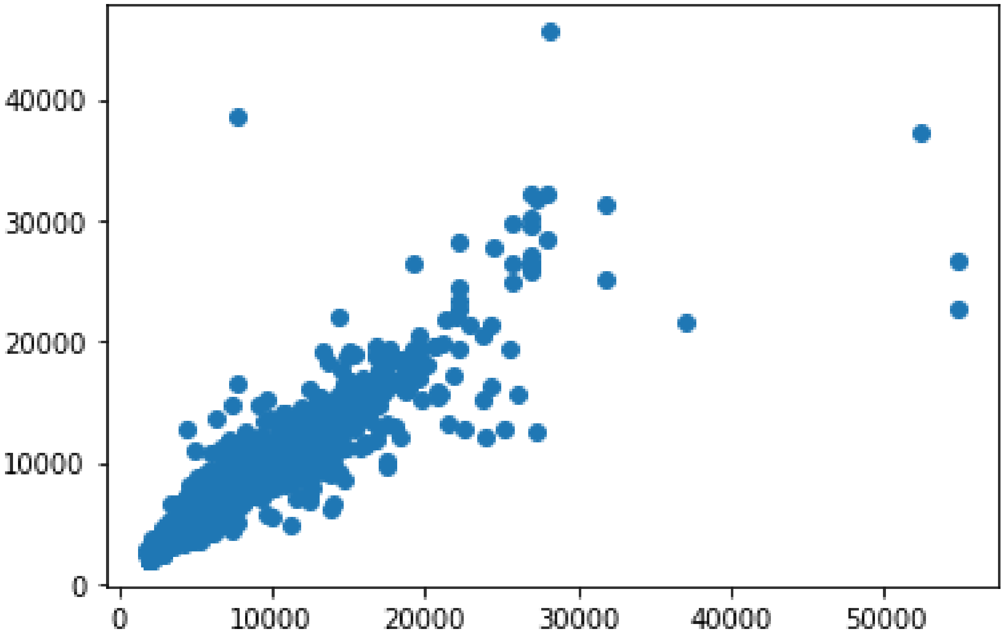
```
[218]: import seaborn as sns

       sns.distplot(y_test-y_pred)
```

```
[218]: <matplotlib.axes._subplots.AxesSubplot at 0x20ea3c6e508>
```



[220]:
```
plt.scatter(y_test,y_pred)
```

[220]:
```
<matplotlib.collections.PathCollection at 0x20ea2b75d08>
```



[225]:

```
from sklearn import metrics

print('MAE:', metrics.mean_absolute_error(y_test, y_pred))
print('MSE:', metrics.mean_squared_error(y_test,  y_pred))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test,  y_pred)))
```

```
MAE: 674.6057706378028
MSE: 2697355.2626544232
RMSE: 1642.3627073988325
```

[ ]:

```

```