bulldozer-price-regression

June 16, 2021

1 Predicting the Sale Price of Bulldozers using Machine Learning

In this notebook, we are going to work with the goal of predicting the sale price of bulldozers.

1.1 Problem Definition

> How well can we predict the future sale price of a bulldozer, given its characteristics and for how much similar bulldozers previously have been sold for?

1.2 Data

The data is downloaded from Kaggle Bluebook for Bulldozers competition: https://www.kaggle.com/c/bluebook-for-bulldozers/data

There are three main datasets:

- Train.csv is the training set, which contains data through the end of 2011.
- Valid.csv is the validation set, which contains data from January 1, 2012 - April 30, 2012 You make predictions on this set throughout the majority of the competition. Your score on this set is used to create the public leaderboard.
- Test.csv is the test set, which won't be released until the last week of the competition. It contains data from May 1, 2012 - November 2012. Your score on the test set determines your final rank for the competition.

1.3 Evaluation

The evaluation metric for this project is the RMSLE (root mean squared log error) between the actual and predicted auction prices.

Note: The goal for most regression evaluation metrics is to minimize the error. For example, our goal for this project will be to buld a machine learning model that minimizes RMSLE

1.4 Features

Kaggle provides a data dictionary which has detalling of all the features of the dataset. You can find the data dictionary here: https://drive.google.com/file/d/1S7bTkXFfFOlm388K9z-Z41Y3LHyUR0TR/view?usp=sharing

[1]:
```
# Importing the tools for the project

import pandas as pd
```

```
import numpy as np
import matplotlib.pyplot as plt
import sklearn

%matplotlib inline
```

[2]:
```
# Importing train and validation sets

df = pd.read_csv("data/TrainAndValid.csv", low_memory=False)
```

[3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 412698 entries, 0 to 412697
Data columns (total 53 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   SalesID                   412698 non-null  int64
 1   SalePrice                 412698 non-null  float64
 2   MachineID                 412698 non-null  int64
 3   ModelID                   412698 non-null  int64
 4   datasource                412698 non-null  int64
 5   auctioneerID              392562 non-null  float64
 6   YearMade                  412698 non-null  int64
 7   MachineHoursCurrentMeter  147504 non-null  float64
 8   UsageBand                 73670 non-null   object
 9   saledate                  412698 non-null  object
 10  fiModelDesc               412698 non-null  object
 11  fiBaseModel               412698 non-null  object
 12  fiSecondaryDesc           271971 non-null  object
 13  fiModelSeries             58667 non-null   object
 14  fiModelDescriptor         74816 non-null   object
 15  ProductSize               196093 non-null  object
 16  fiProductClassDesc        412698 non-null  object
 17  state                     412698 non-null  object
 18  ProductGroup              412698 non-null  object
 19  ProductGroupDesc          412698 non-null  object
 20  Drive_System              107087 non-null  object
 21  Enclosure                 412364 non-null  object
 22  Forks                     197715 non-null  object
 23  Pad_Type                  81096 non-null   object
 24  Ride_Control              152728 non-null  object
 25  Stick                     81096 non-null   object
 26  Transmission              188007 non-null  object
 27  Turbocharged              81096 non-null   object
 28  Blade_Extension           25983 non-null   object
 29  Blade_Width               25983 non-null   object
 30  Enclosure_Type            25983 non-null   object
```

```
 31  Engine_Horsepower         25983 non-null   object
 32  Hydraulics                330133 non-null  object
 33  Pushblock                 25983 non-null   object
 34  Ripper                    106945 non-null  object
 35  Scarifier                 25994 non-null   object
 36  Tip_Control               25983 non-null   object
 37  Tire_Size                 97638 non-null   object
 38  Coupler                   220679 non-null  object
 39  Coupler_System            44974 non-null   object
 40  Grouser_Tracks            44875 non-null   object
 41  Hydraulics_Flow           44875 non-null   object
 42  Track_Type                102193 non-null  object
 43  Undercarriage_Pad_Width   102916 non-null  object
 44  Stick_Length              102261 non-null  object
 45  Thumb                     102332 non-null  object
 46  Pattern_Changer           102261 non-null  object
 47  Grouser_Type              102193 non-null  object
 48  Backhoe_Mounting          80712 non-null   object
 49  Blade_Type                81875 non-null   object
 50  Travel_Controls           81877 non-null   object
 51  Differential_Type         71564 non-null   object
 52  Steering_Controls         71522 non-null   object
dtypes: float64(3), int64(5), object(45)
memory usage: 166.9+ MB
```

[4]:
```
# let's see how many missing values each column have
df.isna().sum()
```

[4]:
```
SalesID                         0
SalePrice                       0
MachineID                       0
ModelID                         0
datasource                      0
auctioneerID                20136
YearMade                        0
MachineHoursCurrentMeter    265194
UsageBand                   339028
saledate                        0
fiModelDesc                     0
fiBaseModel                     0
fiSecondaryDesc            140727
fiModelSeries              354031
fiModelDescriptor          337882
ProductSize                216605
fiProductClassDesc              0
state                           0
ProductGroup                    0
```

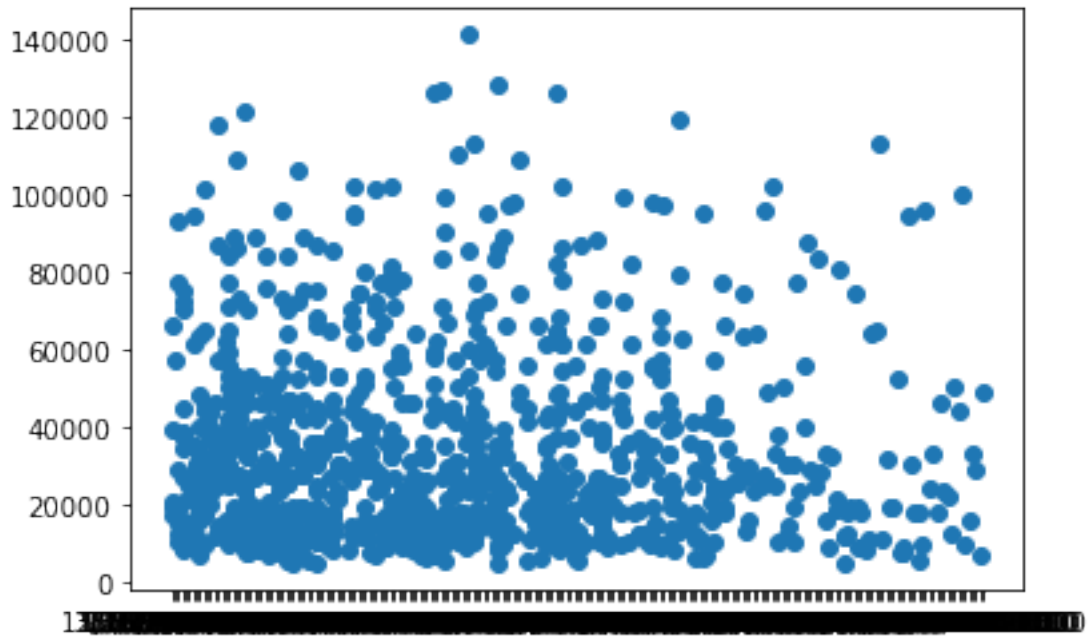
```
ProductGroupDesc              0
Drive_System             305611
Enclosure                   334
Forks                    214983
Pad_Type                 331602
Ride_Control             259970
Stick                    331602
Transmission             224691
Turbocharged             331602
Blade_Extension          386715
Blade_Width              386715
Enclosure_Type           386715
Engine_Horsepower        386715
Hydraulics                82565
Pushblock                386715
Ripper                   305753
Scarifier                386704
Tip_Control              386715
Tire_Size                315060
Coupler                  192019
Coupler_System           367724
Grouser_Tracks           367823
Hydraulics_Flow          367823
Track_Type               310505
Undercarriage_Pad_Width  309782
Stick_Length             310437
Thumb                    310366
Pattern_Changer          310437
Grouser_Type             310505
Backhoe_Mounting         331986
Blade_Type               330823
Travel_Controls          330821
Differential_Type        341134
Steering_Controls        341176
dtype: int64
```

[5]:
```
# Let's see the column names
df.columns
```

[5]:
```
Index(['SalesID', 'SalePrice', 'MachineID', 'ModelID', 'datasource',
       'auctioneerID', 'YearMade', 'MachineHoursCurrentMeter', 'UsageBand',
       'saledate', 'fiModelDesc', 'fiBaseModel', 'fiSecondaryDesc',
       'fiModelSeries', 'fiModelDescriptor', 'ProductSize',
       'fiProductClassDesc', 'state', 'ProductGroup', 'ProductGroupDesc',
       'Drive_System', 'Enclosure', 'Forks', 'Pad_Type', 'Ride_Control',
       'Stick', 'Transmission', 'Turbocharged', 'Blade_Extension',
       'Blade_Width', 'Enclosure_Type', 'Engine_Horsepower', 'Hydraulics',
```

```
       'Pushblock', 'Ripper', 'Scarifier', 'Tip_Control', 'Tire_Size',
       'Coupler', 'Coupler_System', 'Grouser_Tracks', 'Hydraulics_Flow',
       'Track_Type', 'Undercarriage_Pad_Width', 'Stick_Length', 'Thumb',
       'Pattern_Changer', 'Grouser_Type', 'Backhoe_Mounting', 'Blade_Type',
       'Travel_Controls', 'Differential_Type', 'Steering_Controls'],
      dtype='object')
```

[6]:
```
fig, ax = plt.subplots()
ax.scatter(df["saledate"][:1000], df["SalePrice"][:1000]);
```



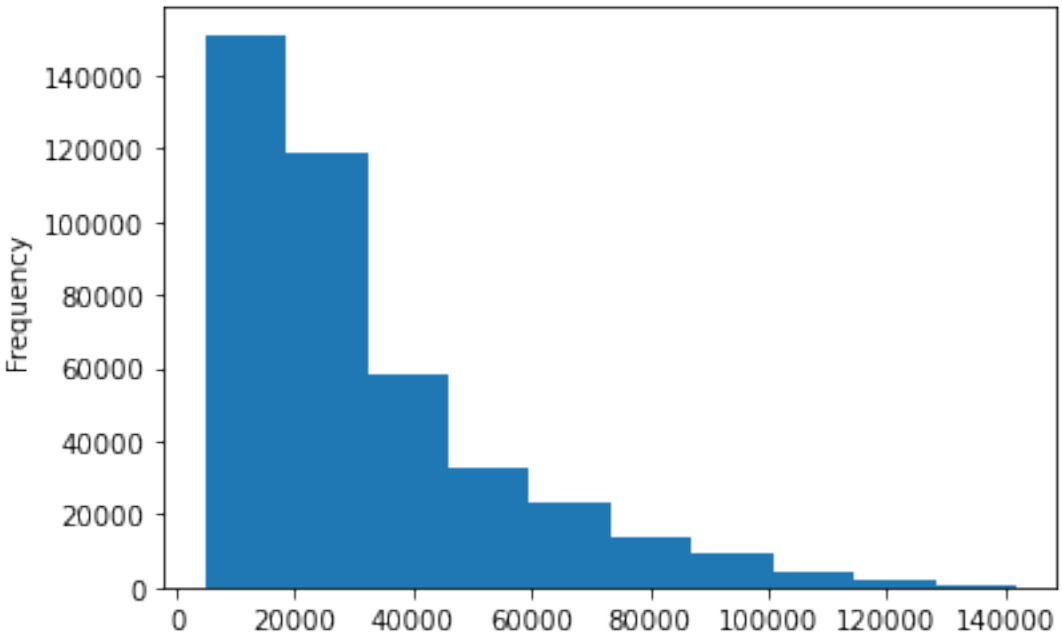
[7]:
```
df["saledate"][:1000]
```

[7]:
```
0      11/16/2006 0:00
1       3/26/2004 0:00
2       2/26/2004 0:00
3       5/19/2011 0:00
4       7/23/2009 0:00
             ...
995     7/16/2009 0:00
996     6/14/2007 0:00
997     9/22/2005 0:00
998     7/28/2005 0:00
999     6/16/2011 0:00
Name: saledate, Length: 1000, dtype: object
```

[8]:
```
df["saledate"].dtype
```

[8]:
```
dtype('O')
```

[9]:
```
df["SalePrice"].plot.hist();
```



1.4.1 Parsing Dates

When we work with time series data, we want to enrich the time and date component as much as possible.

We can do that by telling pandas which column in our dataset contains the date using the `parse_dates` parameter.

[10]:
```
# Importing data again, but this time with parse_dates parameter

df = pd.read_csv("data/TrainAndValid.csv",
                 low_memory=False,
                 parse_dates=["saledate"])
```

[11]:
```
df["saledate"].dtype
```

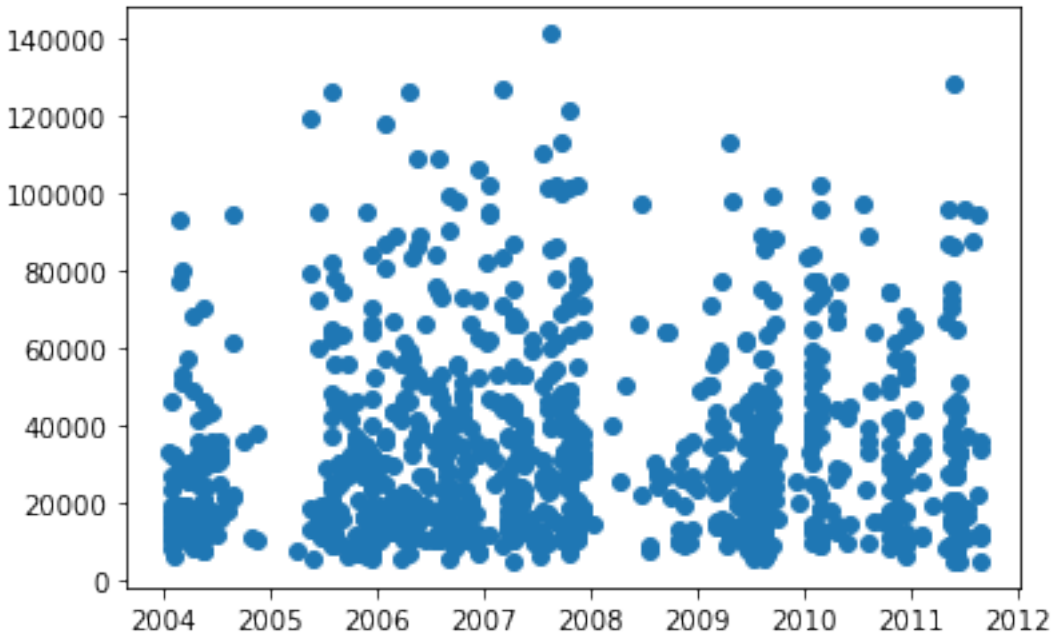
[11]:
```
dtype('<M8[ns]')
```

[12]:
```
df["saledate"][:1000]
```

[12]:
```
0      2006-11-16
1      2004-03-26
2      2004-02-26
```

```
3     2011-05-19
4     2009-07-23
         …
995   2009-07-16
996   2007-06-14
997   2005-09-22
998   2005-07-28
999   2011-06-16
Name: saledate, Length: 1000, dtype: datetime64[ns]
```

[13]:
```
fig, ax = plt.subplots()
ax.scatter(df["saledate"][:1000], df["SalePrice"][:1000]);
```



[14]:
```
df.head()
```

[14]:

	SalesID	SalePrice	MachineID	ModelID	datasource	auctioneerID	YearMade	\
0	1139246	66000.0	999089	3157	121	3.0	2004	
1	1139248	57000.0	117657	77	121	3.0	1996	
2	1139249	10000.0	434808	7009	121	3.0	2001	
3	1139251	38500.0	1026470	332	121	3.0	2001	
4	1139253	11000.0	1057373	17311	121	3.0	2007	

	MachineHoursCurrentMeter	UsageBand	saledate	…	Undercarriage_Pad_Width	\
0	68.0	Low	2006-11-16	…	NaN	
1	4640.0	Low	2004-03-26	…	NaN	

```
2                    2838.0      High 2004-02-26  …                    NaN
3                    3486.0      High 2011-05-19  …                    NaN
4                     722.0    Medium 2009-07-23  …                    NaN

  Stick_Length Thumb Pattern_Changer Grouser_Type Backhoe_Mounting Blade_Type  \
0          NaN   NaN             NaN          NaN              NaN         NaN
1          NaN   NaN             NaN          NaN              NaN         NaN
2          NaN   NaN             NaN          NaN              NaN         NaN
3          NaN   NaN             NaN          NaN              NaN         NaN
4          NaN   NaN             NaN          NaN              NaN         NaN

  Travel_Controls Differential_Type Steering_Controls
0             NaN          Standard      Conventional
1             NaN          Standard      Conventional
2             NaN               NaN               NaN
3             NaN               NaN               NaN
4             NaN               NaN               NaN

[5 rows x 53 columns]
```

[15]:
```
df.head().T
```

[15]:
```
                                                          0  \
SalesID                                             1139246
SalePrice                                           66000.0
MachineID                                            999089
ModelID                                                3157
datasource                                              121
auctioneerID                                            3.0
YearMade                                               2004
MachineHoursCurrentMeter                               68.0
UsageBand                                               Low
saledate                                2006-11-16 00:00:00
fiModelDesc                                            521D
fiBaseModel                                             521
fiSecondaryDesc                                           D
fiModelSeries                                           NaN
fiModelDescriptor                                       NaN
ProductSize                                             NaN
fiProductClassDesc        Wheel Loader - 110.0 to 120.0 Horsepower
state                                                Alabama
ProductGroup                                              WL
ProductGroupDesc                                Wheel Loader
Drive_System                                            NaN
Enclosure                                         EROPS w AC
Forks                                   None or Unspecified
Pad_Type                                                NaN
```

Ride_Control	None or Unspecified
Stick	NaN
Transmission	NaN
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	NaN
Scarifier	NaN
Tip_Control	NaN
Tire_Size	None or Unspecified
Coupler	None or Unspecified
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	NaN
Blade_Type	NaN
Travel_Controls	NaN
Differential_Type	Standard
Steering_Controls	Conventional

	1 \
SalesID	1139248
SalePrice	57000.0
MachineID	117657
ModelID	77
datasource	121
auctioneerID	3.0
YearMade	1996
MachineHoursCurrentMeter	4640.0
UsageBand	Low
saledate	2004-03-26 00:00:00
fiModelDesc	950FII
fiBaseModel	950
fiSecondaryDesc	F
fiModelSeries	II
fiModelDescriptor	NaN
ProductSize	Medium

fiProductClassDesc	Wheel Loader - 150.0 to 175.0 Horsepower
state	North Carolina
ProductGroup	WL
ProductGroupDesc	Wheel Loader
Drive_System	NaN
Enclosure	EROPS w AC
Forks	None or Unspecified
Pad_Type	NaN
Ride_Control	None or Unspecified
Stick	NaN
Transmission	NaN
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	NaN
Scarifier	NaN
Tip_Control	NaN
Tire_Size	23.5
Coupler	None or Unspecified
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	NaN
Blade_Type	NaN
Travel_Controls	NaN
Differential_Type	Standard
Steering_Controls	Conventional

	2 \
SalesID	1139249
SalePrice	10000.0
MachineID	434808
ModelID	7009
datasource	121
auctioneerID	3.0
YearMade	2001
MachineHoursCurrentMeter	2838.0

```
UsageBand                                                        High
saledate                                              2004-02-26 00:00:00
fiModelDesc                                                       226
fiBaseModel                                                       226
fiSecondaryDesc                                                   NaN
fiModelSeries                                                     NaN
fiModelDescriptor                                                 NaN
ProductSize                                                       NaN
fiProductClassDesc       Skid Steer Loader - 1351.0 to 1601.0 Lb Operat…
state                                                        New York
ProductGroup                                                      SSL
ProductGroupDesc                                   Skid Steer Loaders
Drive_System                                                      NaN
Enclosure                                                      OROPS
Forks                                             None or Unspecified
Pad_Type                                                          NaN
Ride_Control                                                      NaN
Stick                                                             NaN
Transmission                                                      NaN
Turbocharged                                                      NaN
Blade_Extension                                                   NaN
Blade_Width                                                       NaN
Enclosure_Type                                                    NaN
Engine_Horsepower                                                 NaN
Hydraulics                                                  Auxiliary
Pushblock                                                         NaN
Ripper                                                            NaN
Scarifier                                                         NaN
Tip_Control                                                       NaN
Tire_Size                                                         NaN
Coupler                                           None or Unspecified
Coupler_System                                    None or Unspecified
Grouser_Tracks                                    None or Unspecified
Hydraulics_Flow                                          Standard
Track_Type                                                        NaN
Undercarriage_Pad_Width                                           NaN
Stick_Length                                                      NaN
Thumb                                                             NaN
Pattern_Changer                                                   NaN
Grouser_Type                                                      NaN
Backhoe_Mounting                                                  NaN
Blade_Type                                                        NaN
Travel_Controls                                                   NaN
Differential_Type                                                 NaN
Steering_Controls                                                 NaN

                                                                3  \
```

SalesID	1139251
SalePrice	38500.0
MachineID	1026470
ModelID	332
datasource	121
auctioneerID	3.0
YearMade	2001
MachineHoursCurrentMeter	3486.0
UsageBand	High
saledate	2011-05-19 00:00:00
fiModelDesc	PC120-6E
fiBaseModel	PC120
fiSecondaryDesc	NaN
fiModelSeries	-6E
fiModelDescriptor	NaN
ProductSize	Small
fiProductClassDesc	Hydraulic Excavator, Track - 12.0 to 14.0 Metr…
state	Texas
ProductGroup	TEX
ProductGroupDesc	Track Excavators
Drive_System	NaN
Enclosure	EROPS w AC
Forks	NaN
Pad_Type	NaN
Ride_Control	NaN
Stick	NaN
Transmission	NaN
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	NaN
Scarifier	NaN
Tip_Control	NaN
Tire_Size	NaN
Coupler	None or Unspecified
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN

Grouser_Type	NaN
Backhoe_Mounting	NaN
Blade_Type	NaN
Travel_Controls	NaN
Differential_Type	NaN
Steering_Controls	NaN
	4
SalesID	1139253
SalePrice	11000.0
MachineID	1057373
ModelID	17311
datasource	121
auctioneerID	3.0
YearMade	2007
MachineHoursCurrentMeter	722.0
UsageBand	Medium
saledate	2009-07-23 00:00:00
fiModelDesc	S175
fiBaseModel	S175
fiSecondaryDesc	NaN
fiModelSeries	NaN
fiModelDescriptor	NaN
ProductSize	NaN
fiProductClassDesc	Skid Steer Loader - 1601.0 to 1751.0 Lb Operat…
state	New York
ProductGroup	SSL
ProductGroupDesc	Skid Steer Loaders
Drive_System	NaN
Enclosure	EROPS
Forks	None or Unspecified
Pad_Type	NaN
Ride_Control	NaN
Stick	NaN
Transmission	NaN
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	Auxiliary
Pushblock	NaN
Ripper	NaN
Scarifier	NaN
Tip_Control	NaN
Tire_Size	NaN
Coupler	None or Unspecified

```
Coupler_System                                  None or Unspecified
Grouser_Tracks                                  None or Unspecified
Hydraulics_Flow                                            Standard
Track_Type                                                      NaN
Undercarriage_Pad_Width                                         NaN
Stick_Length                                                    NaN
Thumb                                                           NaN
Pattern_Changer                                                 NaN
Grouser_Type                                                    NaN
Backhoe_Mounting                                                NaN
Blade_Type                                                      NaN
Travel_Controls                                                 NaN
Differential_Type                                               NaN
Steering_Controls                                               NaN
```

[16]:
```
df["saledate"].head(20)
```

[16]:
```
0    2006-11-16
1    2004-03-26
2    2004-02-26
3    2011-05-19
4    2009-07-23
5    2008-12-18
6    2004-08-26
7    2005-11-17
8    2009-08-27
9    2007-08-09
10   2008-08-21
11   2006-08-24
12   2005-10-20
13   2006-01-26
14   2006-01-03
15   2006-11-16
16   2007-06-14
17   2010-01-28
18   2006-03-09
19   2005-11-17
Name: saledate, dtype: datetime64[ns]
```

1.4.2 Sort the dataset by saledate

We can see the the dates are not in order. If we sort the dataset by date, we might get a more clear idea abour the dataset.

When working with timeseries data, it's a good idea to sort the data by date.

[17]:
```
# Sort dataframe by saledate
df.sort_values(by=["saledate"], inplace=True, ascending=True)
```

```
df["saledate"].head(20)
```

[17]:
```
205615   1989-01-17
274835   1989-01-31
141296   1989-01-31
212552   1989-01-31
62755    1989-01-31
54653    1989-01-31
81383    1989-01-31
204924   1989-01-31
135376   1989-01-31
113390   1989-01-31
113394   1989-01-31
116419   1989-01-31
32138    1989-01-31
127610   1989-01-31
76171    1989-01-31
127000   1989-01-31
128130   1989-01-31
127626   1989-01-31
55455    1989-01-31
55454    1989-01-31
Name: saledate, dtype: datetime64[ns]
```

1.4.3 Making a copy of the original data

We should keep a copy of the original data.

[18]:
```
# Copy of original data
df_tmp = df.copy()
```

[19]:
```
df_tmp["saledate"].head(20)
```

[19]:
```
205615   1989-01-17
274835   1989-01-31
141296   1989-01-31
212552   1989-01-31
62755    1989-01-31
54653    1989-01-31
81383    1989-01-31
204924   1989-01-31
135376   1989-01-31
113390   1989-01-31
113394   1989-01-31
116419   1989-01-31
32138    1989-01-31
127610   1989-01-31
76171    1989-01-31
```

```
127000   1989-01-31
128130   1989-01-31
127626   1989-01-31
55455    1989-01-31
55454    1989-01-31
Name: saledate, dtype: datetime64[ns]
```

[20]:
```
# Feature engineering out dataset
df_tmp["saleYear"] = df_tmp["saledate"].dt.year
df_tmp["saleMonth"] = df_tmp["saledate"].dt.month
df_tmp["saleDay"] = df_tmp["saledate"].dt.day
df_tmp["saleDayOfWeek"] = df_tmp["saledate"].dt.dayofweek
df_tmp["saleDayOfYear"] = df_tmp["saledate"].dt.dayofyear
```

[21]:
```
df_tmp.head().T
```

[21]:

	205615 \
SalesID	1646770
SalePrice	9500.0
MachineID	1126363
ModelID	8434
datasource	132
auctioneerID	18.0
YearMade	1974
MachineHoursCurrentMeter	NaN
UsageBand	NaN
saledate	1989-01-17 00:00:00
fiModelDesc	TD20
fiBaseModel	TD20
fiSecondaryDesc	NaN
fiModelSeries	NaN
fiModelDescriptor	NaN
ProductSize	Medium
fiProductClassDesc	Track Type Tractor, Dozer - 105.0 to 130.0 Hor…
state	Texas
ProductGroup	TTT
ProductGroupDesc	Track Type Tractors
Drive_System	NaN
Enclosure	OROPS
Forks	NaN
Pad_Type	NaN
Ride_Control	NaN
Stick	NaN
Transmission	Direct Drive
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN

Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	None or Unspecified
Scarifier	NaN
Tip_Control	NaN
Tire_Size	NaN
Coupler	NaN
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	None or Unspecified
Blade_Type	Straight
Travel_Controls	None or Unspecified
Differential_Type	NaN
Steering_Controls	NaN
saleYear	1989
saleMonth	1
saleDay	17
saleDayOfWeek	1
saleDayOfYear	17

	274835 \
SalesID	1821514
SalePrice	14000.0
MachineID	1194089
ModelID	10150
datasource	132
auctioneerID	99.0
YearMade	1980
MachineHoursCurrentMeter	NaN
UsageBand	NaN
saledate	1989-01-31 00:00:00
fiModelDesc	A66
fiBaseModel	A66
fiSecondaryDesc	NaN
fiModelSeries	NaN
fiModelDescriptor	NaN
ProductSize	NaN
fiProductClassDesc	Wheel Loader - 120.0 to 135.0 Horsepower

state	Florida
ProductGroup	WL
ProductGroupDesc	Wheel Loader
Drive_System	NaN
Enclosure	OROPS
Forks	None or Unspecified
Pad_Type	NaN
Ride_Control	None or Unspecified
Stick	NaN
Transmission	NaN
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	NaN
Scarifier	NaN
Tip_Control	NaN
Tire_Size	None or Unspecified
Coupler	None or Unspecified
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	NaN
Blade_Type	NaN
Travel_Controls	NaN
Differential_Type	Standard
Steering_Controls	Conventional
saleYear	1989
saleMonth	1
saleDay	31
saleDayOfWeek	1
saleDayOfYear	31

	141296 \
SalesID	1505138
SalePrice	50000.0
MachineID	1473654
ModelID	4139

datasource	132
auctioneerID	99.0
YearMade	1978
MachineHoursCurrentMeter	NaN
UsageBand	NaN
saledate	1989-01-31 00:00:00
fiModelDesc	D7G
fiBaseModel	D7
fiSecondaryDesc	G
fiModelSeries	NaN
fiModelDescriptor	NaN
ProductSize	Large
fiProductClassDesc	Track Type Tractor, Dozer - 190.0 to 260.0 Hor…
state	Florida
ProductGroup	TTT
ProductGroupDesc	Track Type Tractors
Drive_System	NaN
Enclosure	OROPS
Forks	NaN
Pad_Type	NaN
Ride_Control	NaN
Stick	NaN
Transmission	Standard
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	None or Unspecified
Scarifier	NaN
Tip_Control	NaN
Tire_Size	NaN
Coupler	NaN
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	None or Unspecified
Blade_Type	Straight
Travel_Controls	None or Unspecified

Differential_Type	NaN
Steering_Controls	NaN
saleYear	1989
saleMonth	1
saleDay	31
saleDayOfWeek	1
saleDayOfYear	31

	212552 \
SalesID	1671174
SalePrice	16000.0
MachineID	1327630
ModelID	8591
datasource	132
auctioneerID	99.0
YearMade	1980
MachineHoursCurrentMeter	NaN
UsageBand	NaN
saledate	1989-01-31 00:00:00
fiModelDesc	A62
fiBaseModel	A62
fiSecondaryDesc	NaN
fiModelSeries	NaN
fiModelDescriptor	NaN
ProductSize	NaN
fiProductClassDesc	Wheel Loader - Unidentified
state	Florida
ProductGroup	WL
ProductGroupDesc	Wheel Loader
Drive_System	NaN
Enclosure	EROPS
Forks	None or Unspecified
Pad_Type	NaN
Ride_Control	None or Unspecified
Stick	NaN
Transmission	NaN
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	NaN
Scarifier	NaN
Tip_Control	NaN
Tire_Size	None or Unspecified

Coupler	None or Unspecified
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	NaN
Blade_Type	NaN
Travel_Controls	NaN
Differential_Type	Standard
Steering_Controls	Conventional
saleYear	1989
saleMonth	1
saleDay	31
saleDayOfWeek	1
saleDayOfYear	31

	62755
SalesID	1329056
SalePrice	22000.0
MachineID	1336053
ModelID	4089
datasource	132
auctioneerID	99.0
YearMade	1984
MachineHoursCurrentMeter	NaN
UsageBand	NaN
saledate	1989-01-31 00:00:00
fiModelDesc	D3B
fiBaseModel	D3
fiSecondaryDesc	B
fiModelSeries	NaN
fiModelDescriptor	NaN
ProductSize	NaN
fiProductClassDesc	Track Type Tractor, Dozer - 20.0 to 75.0 Horse…
state	Florida
ProductGroup	TTT
ProductGroupDesc	Track Type Tractors
Drive_System	NaN
Enclosure	OROPS
Forks	NaN
Pad_Type	NaN
Ride_Control	NaN

```
Stick                                                  NaN
Transmission                                      Standard
Turbocharged                                           NaN
Blade_Extension                                        NaN
Blade_Width                                            NaN
Enclosure_Type                                         NaN
Engine_Horsepower                                      NaN
Hydraulics                                         2 Valve
Pushblock                                              NaN
Ripper                                  None or Unspecified
Scarifier                                              NaN
Tip_Control                                            NaN
Tire_Size                                              NaN
Coupler                                                NaN
Coupler_System                                         NaN
Grouser_Tracks                                         NaN
Hydraulics_Flow                                        NaN
Track_Type                                             NaN
Undercarriage_Pad_Width                                NaN
Stick_Length                                           NaN
Thumb                                                  NaN
Pattern_Changer                                        NaN
Grouser_Type                                           NaN
Backhoe_Mounting                        None or Unspecified
Blade_Type                                             PAT
Travel_Controls                                      Lever
Differential_Type                                      NaN
Steering_Controls                                      NaN
saleYear                                              1989
saleMonth                                                1
saleDay                                                 31
saleDayOfWeek                                            1
saleDayOfYear                                           31
```

[22]:
```
# As we have enriched our dataframe with date time features, we can now drop␣
 ↪the sale date column
df_tmp.drop("saledate", axis=1, inplace=True)
```

[23]:
```
df_tmp["state"].value_counts()
```

[23]:
```
Florida          67320
Texas            53110
California       29761
Washington       16222
Georgia          14633
Maryland         13322
Mississippi      13240
```

```
Ohio              12369
Illinois          11540
Colorado          11529
New Jersey        11156
North Carolina    10636
Tennessee         10298
Alabama           10292
Pennsylvania      10234
South Carolina     9951
Arizona            9364
New York           8639
Connecticut        8276
Minnesota          7885
Missouri           7178
Nevada             6932
Louisiana          6627
Kentucky           5351
Maine              5096
Indiana            4124
Arkansas           3933
New Mexico         3631
Utah               3046
Unspecified        2801
Wisconsin          2745
New Hampshire      2738
Virginia           2353
Idaho              2025
Oregon             1911
Michigan           1831
Wyoming            1672
Iowa               1336
Montana            1336
Oklahoma           1326
Nebraska            866
West Virginia       840
Kansas              667
Delaware            510
North Dakota        480
Alaska              430
Massachusetts       347
Vermont             300
South Dakota        244
Hawaii              118
Rhode Island         83
Puerto Rico          42
Washington DC         2
Name: state, dtype: int64
```

1.5 Modelling

[24]:
```
# let's try and fit a model to our dataset
# from sklearn.ensemble import RandomForestRegressor

# model = RandomForestRegressor(n_jobs=-1,
#                               random_state=42)

# model.fit(df_tmp.drop("SalePrice", axis=1), df_tmp["SalePrice"])
```

[25]:
```
df_tmp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 412698 entries, 205615 to 409203
Data columns (total 57 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   SalesID                   412698 non-null  int64
 1   SalePrice                 412698 non-null  float64
 2   MachineID                 412698 non-null  int64
 3   ModelID                   412698 non-null  int64
 4   datasource                412698 non-null  int64
 5   auctioneerID              392562 non-null  float64
 6   YearMade                  412698 non-null  int64
 7   MachineHoursCurrentMeter  147504 non-null  float64
 8   UsageBand                 73670 non-null   object
 9   fiModelDesc               412698 non-null  object
 10  fiBaseModel               412698 non-null  object
 11  fiSecondaryDesc           271971 non-null  object
 12  fiModelSeries             58667 non-null   object
 13  fiModelDescriptor         74816 non-null   object
 14  ProductSize               196093 non-null  object
 15  fiProductClassDesc        412698 non-null  object
 16  state                     412698 non-null  object
 17  ProductGroup              412698 non-null  object
 18  ProductGroupDesc          412698 non-null  object
 19  Drive_System              107087 non-null  object
 20  Enclosure                 412364 non-null  object
 21  Forks                     197715 non-null  object
 22  Pad_Type                  81096 non-null   object
 23  Ride_Control              152728 non-null  object
 24  Stick                     81096 non-null   object
 25  Transmission              188007 non-null  object
 26  Turbocharged              81096 non-null   object
 27  Blade_Extension           25983 non-null   object
 28  Blade_Width               25983 non-null   object
 29  Enclosure_Type            25983 non-null   object
 30  Engine_Horsepower         25983 non-null   object
```

```
 31  Hydraulics                330133 non-null  object
 32  Pushblock                 25983 non-null   object
 33  Ripper                    106945 non-null  object
 34  Scarifier                 25994 non-null   object
 35  Tip_Control               25983 non-null   object
 36  Tire_Size                 97638 non-null   object
 37  Coupler                   220679 non-null  object
 38  Coupler_System            44974 non-null   object
 39  Grouser_Tracks            44875 non-null   object
 40  Hydraulics_Flow           44875 non-null   object
 41  Track_Type                102193 non-null  object
 42  Undercarriage_Pad_Width   102916 non-null  object
 43  Stick_Length              102261 non-null  object
 44  Thumb                     102332 non-null  object
 45  Pattern_Changer           102261 non-null  object
 46  Grouser_Type              102193 non-null  object
 47  Backhoe_Mounting          80712 non-null   object
 48  Blade_Type                81875 non-null   object
 49  Travel_Controls           81877 non-null   object
 50  Differential_Type         71564 non-null   object
 51  Steering_Controls         71522 non-null   object
 52  saleYear                  412698 non-null  int64
 53  saleMonth                 412698 non-null  int64
 54  saleDay                   412698 non-null  int64
 55  saleDayOfWeek             412698 non-null  int64
 56  saleDayOfYear             412698 non-null  int64
dtypes: float64(3), int64(10), object(44)
memory usage: 182.6+ MB
```

As we can see we have a lot of missing and non-numeric data. Before trying to fit all these into a model, we have to take care of these values

[26]:
```
pd.api.types.is_string_dtype(df["UsageBand"])
```

[26]:
```
True
```

[27]:
```
# Find the columns that contain string
for label, content in df_tmp.items():
    if pd.api.types.is_string_dtype(content):
        print(label)
```

```
UsageBand
fiModelDesc
fiBaseModel
fiSecondaryDesc
fiModelSeries
fiModelDescriptor
ProductSize
fiProductClassDesc
```

```
state
ProductGroup
ProductGroupDesc
Drive_System
Enclosure
Forks
Pad_Type
Ride_Control
Stick
Transmission
Turbocharged
Blade_Extension
Blade_Width
Enclosure_Type
Engine_Horsepower
Hydraulics
Pushblock
Ripper
Scarifier
Tip_Control
Tire_Size
Coupler
Coupler_System
Grouser_Tracks
Hydraulics_Flow
Track_Type
Undercarriage_Pad_Width
Stick_Length
Thumb
Pattern_Changer
Grouser_Type
Backhoe_Mounting
Blade_Type
Travel_Controls
Differential_Type
Steering_Controls
```

[28]:
```
# Now let's turn all the string values into categorical values
for label, content in df_tmp.items():
    if pd.api.types.is_string_dtype(content):
        df_tmp[label] = content.astype("category").cat.as_ordered()
```

[29]:
```
df_tmp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 412698 entries, 205615 to 409203
Data columns (total 57 columns):
 #   Column                    Non-Null Count   Dtype
```

```
---  ------                    --------------   -----
 0   SalesID                   412698 non-null  int64
 1   SalePrice                 412698 non-null  float64
 2   MachineID                 412698 non-null  int64
 3   ModelID                   412698 non-null  int64
 4   datasource                412698 non-null  int64
 5   auctioneerID              392562 non-null  float64
 6   YearMade                  412698 non-null  int64
 7   MachineHoursCurrentMeter  147504 non-null  float64
 8   UsageBand                 73670 non-null   category
 9   fiModelDesc               412698 non-null  category
 10  fiBaseModel               412698 non-null  category
 11  fiSecondaryDesc           271971 non-null  category
 12  fiModelSeries             58667 non-null   category
 13  fiModelDescriptor         74816 non-null   category
 14  ProductSize               196093 non-null  category
 15  fiProductClassDesc        412698 non-null  category
 16  state                     412698 non-null  category
 17  ProductGroup              412698 non-null  category
 18  ProductGroupDesc          412698 non-null  category
 19  Drive_System              107087 non-null  category
 20  Enclosure                 412364 non-null  category
 21  Forks                     197715 non-null  category
 22  Pad_Type                  81096 non-null   category
 23  Ride_Control              152728 non-null  category
 24  Stick                     81096 non-null   category
 25  Transmission              188007 non-null  category
 26  Turbocharged              81096 non-null   category
 27  Blade_Extension           25983 non-null   category
 28  Blade_Width               25983 non-null   category
 29  Enclosure_Type            25983 non-null   category
 30  Engine_Horsepower         25983 non-null   category
 31  Hydraulics                330133 non-null  category
 32  Pushblock                 25983 non-null   category
 33  Ripper                    106945 non-null  category
 34  Scarifier                 25994 non-null   category
 35  Tip_Control               25983 non-null   category
 36  Tire_Size                 97638 non-null   category
 37  Coupler                   220679 non-null  category
 38  Coupler_System            44974 non-null   category
 39  Grouser_Tracks            44875 non-null   category
 40  Hydraulics_Flow           44875 non-null   category
 41  Track_Type                102193 non-null  category
 42  Undercarriage_Pad_Width   102916 non-null  category
 43  Stick_Length              102261 non-null  category
 44  Thumb                     102332 non-null  category
 45  Pattern_Changer           102261 non-null  category
 46  Grouser_Type              102193 non-null  category
```

```
 47  Backhoe_Mounting          80712 non-null   category
 48  Blade_Type                81875 non-null   category
 49  Travel_Controls           81877 non-null   category
 50  Differential_Type         71564 non-null   category
 51  Steering_Controls         71522 non-null   category
 52  saleYear                  412698 non-null  int64
 53  saleMonth                 412698 non-null  int64
 54  saleDay                   412698 non-null  int64
 55  saleDayOfWeek             412698 non-null  int64
 56  saleDayOfYear             412698 non-null  int64
dtypes: category(44), float64(3), int64(10)
memory usage: 63.2 MB
```

[30]:
```
df_tmp["state"].cat.categories
```

[30]:
```
Index(['Alabama', 'Alaska', 'Arizona', 'Arkansas', 'California', 'Colorado',
       'Connecticut', 'Delaware', 'Florida', 'Georgia', 'Hawaii', 'Idaho',
       'Illinois', 'Indiana', 'Iowa', 'Kansas', 'Kentucky', 'Louisiana',
       'Maine', 'Maryland', 'Massachusetts', 'Michigan', 'Minnesota',
       'Mississippi', 'Missouri', 'Montana', 'Nebraska', 'Nevada',
       'New Hampshire', 'New Jersey', 'New Mexico', 'New York',
       'North Carolina', 'North Dakota', 'Ohio', 'Oklahoma', 'Oregon',
       'Pennsylvania', 'Puerto Rico', 'Rhode Island', 'South Carolina',
       'South Dakota', 'Tennessee', 'Texas', 'Unspecified', 'Utah', 'Vermont',
       'Virginia', 'Washington', 'Washington DC', 'West Virginia', 'Wisconsin',
       'Wyoming'],
      dtype='object')
```

[31]:
```
df_tmp["state"].cat.codes
```

[31]:
```
205615    43
274835     8
141296     8
212552     8
62755      8
          ..
410879     4
412476     4
411927     4
407124     4
409203     4
Length: 412698, dtype: int8
```

With the help of pandas `cat_as_ordered()` method, we have transformed all string values into categorical values.

But we still have missing values in our data. Before trying to fit a model in our data, we have to take care of these missing values.

[32]:
```
# Checking the ratio of values missing in each column

df_tmp.isna().sum()/len(df_tmp)
```

[32]:
```
SalesID                     0.000000
SalePrice                   0.000000
MachineID                   0.000000
ModelID                     0.000000
datasource                  0.000000
auctioneerID                0.048791
YearMade                    0.000000
MachineHoursCurrentMeter    0.642586
UsageBand                   0.821492
fiModelDesc                 0.000000
fiBaseModel                 0.000000
fiSecondaryDesc             0.340993
fiModelSeries               0.857845
fiModelDescriptor           0.818715
ProductSize                 0.524851
fiProductClassDesc          0.000000
state                       0.000000
ProductGroup                0.000000
ProductGroupDesc            0.000000
Drive_System                0.740520
Enclosure                   0.000809
Forks                       0.520921
Pad_Type                    0.803498
Ride_Control                0.629928
Stick                       0.803498
Transmission                0.544444
Turbocharged                0.803498
Blade_Extension             0.937041
Blade_Width                 0.937041
Enclosure_Type              0.937041
Engine_Horsepower           0.937041
Hydraulics                  0.200062
Pushblock                   0.937041
Ripper                      0.740864
Scarifier                   0.937014
Tip_Control                 0.937041
Tire_Size                   0.763415
Coupler                     0.465277
Coupler_System              0.891024
Grouser_Tracks              0.891264
Hydraulics_Flow             0.891264
Track_Type                  0.752378
Undercarriage_Pad_Width     0.750626
```

```
Stick_Length              0.752213
Thumb                     0.752041
Pattern_Changer           0.752213
Grouser_Type              0.752378
Backhoe_Mounting          0.804428
Blade_Type                0.801610
Travel_Controls           0.801606
Differential_Type         0.826595
Steering_Controls         0.826697
saleYear                  0.000000
saleMonth                 0.000000
saleDay                   0.000000
saleDayOfWeek             0.000000
saleDayOfYear             0.000000
dtype: float64
```

1.5.1 Save preprocessed data

[33]:
```
# Exporting our processed data to a csv file

df_tmp.to_csv("data/train-tmp-processed.csv", index=False)
```

[34]:
```
# Import the saved preprocessed data
df_tmp = pd.read_csv("data/train-tmp-processed.csv",
                     low_memory=False)

df_tmp.head().T
```

[34]:
```
                                                                  0  \
SalesID                                                     1646770
SalePrice                                                    9500.0
MachineID                                                   1126363
ModelID                                                        8434
datasource                                                      132
auctioneerID                                                   18.0
YearMade                                                       1974
MachineHoursCurrentMeter                                        NaN
UsageBand                                                       NaN
fiModelDesc                                                    TD20
fiBaseModel                                                    TD20
fiSecondaryDesc                                                 NaN
fiModelSeries                                                   NaN
fiModelDescriptor                                               NaN
ProductSize                                                  Medium
fiProductClassDesc        Track Type Tractor, Dozer - 105.0 to 130.0 Hor…
state                                                        Texas
ProductGroup                                                   TTT
```

ProductGroupDesc	Track Type Tractors
Drive_System	NaN
Enclosure	OROPS
Forks	NaN
Pad_Type	NaN
Ride_Control	NaN
Stick	NaN
Transmission	Direct Drive
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	None or Unspecified
Scarifier	NaN
Tip_Control	NaN
Tire_Size	NaN
Coupler	NaN
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	None or Unspecified
Blade_Type	Straight
Travel_Controls	None or Unspecified
Differential_Type	NaN
Steering_Controls	NaN
saleYear	1989
saleMonth	1
saleDay	17
saleDayOfWeek	1
saleDayOfYear	17

	1 \
SalesID	1821514
SalePrice	14000.0
MachineID	1194089
ModelID	10150
datasource	132
auctioneerID	99.0

YearMade	1980
MachineHoursCurrentMeter	NaN
UsageBand	NaN
fiModelDesc	A66
fiBaseModel	A66
fiSecondaryDesc	NaN
fiModelSeries	NaN
fiModelDescriptor	NaN
ProductSize	NaN
fiProductClassDesc	Wheel Loader - 120.0 to 135.0 Horsepower
state	Florida
ProductGroup	WL
ProductGroupDesc	Wheel Loader
Drive_System	NaN
Enclosure	OROPS
Forks	None or Unspecified
Pad_Type	NaN
Ride_Control	None or Unspecified
Stick	NaN
Transmission	NaN
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	NaN
Scarifier	NaN
Tip_Control	NaN
Tire_Size	None or Unspecified
Coupler	None or Unspecified
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	NaN
Blade_Type	NaN
Travel_Controls	NaN
Differential_Type	Standard
Steering_Controls	Conventional
saleYear	1989

```
saleMonth                                                    1
saleDay                                                     31
saleDayOfWeek                                                1
saleDayOfYear                                               31

                                                             2  \
SalesID                                                1505138
SalePrice                                              50000.0
MachineID                                              1473654
ModelID                                                   4139
datasource                                                 132
auctioneerID                                              99.0
YearMade                                                  1978
MachineHoursCurrentMeter                                   NaN
UsageBand                                                  NaN
fiModelDesc                                                D7G
fiBaseModel                                                 D7
fiSecondaryDesc                                              G
fiModelSeries                                              NaN
fiModelDescriptor                                          NaN
ProductSize                                              Large
fiProductClassDesc        Track Type Tractor, Dozer - 190.0 to 260.0 Hor…
state                                                  Florida
ProductGroup                                               TTT
ProductGroupDesc                           Track Type Tractors
Drive_System                                               NaN
Enclosure                                                OROPS
Forks                                                      NaN
Pad_Type                                                   NaN
Ride_Control                                               NaN
Stick                                                      NaN
Transmission                                          Standard
Turbocharged                                               NaN
Blade_Extension                                            NaN
Blade_Width                                                NaN
Enclosure_Type                                             NaN
Engine_Horsepower                                          NaN
Hydraulics                                           2 Valve
Pushblock                                                  NaN
Ripper                                     None or Unspecified
Scarifier                                                  NaN
Tip_Control                                                NaN
Tire_Size                                                  NaN
Coupler                                                    NaN
Coupler_System                                             NaN
Grouser_Tracks                                             NaN
Hydraulics_Flow                                            NaN
```

Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	None or Unspecified
Blade_Type	Straight
Travel_Controls	None or Unspecified
Differential_Type	NaN
Steering_Controls	NaN
saleYear	1989
saleMonth	1
saleDay	31
saleDayOfWeek	1
saleDayOfYear	31

	3 \
SalesID	1671174
SalePrice	16000.0
MachineID	1327630
ModelID	8591
datasource	132
auctioneerID	99.0
YearMade	1980
MachineHoursCurrentMeter	NaN
UsageBand	NaN
fiModelDesc	A62
fiBaseModel	A62
fiSecondaryDesc	NaN
fiModelSeries	NaN
fiModelDescriptor	NaN
ProductSize	NaN
fiProductClassDesc	Wheel Loader - Unidentified
state	Florida
ProductGroup	WL
ProductGroupDesc	Wheel Loader
Drive_System	NaN
Enclosure	EROPS
Forks	None or Unspecified
Pad_Type	NaN
Ride_Control	None or Unspecified
Stick	NaN
Transmission	NaN
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN

Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	NaN
Scarifier	NaN
Tip_Control	NaN
Tire_Size	None or Unspecified
Coupler	None or Unspecified
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	NaN
Blade_Type	NaN
Travel_Controls	NaN
Differential_Type	Standard
Steering_Controls	Conventional
saleYear	1989
saleMonth	1
saleDay	31
saleDayOfWeek	1
saleDayOfYear	31

	4
SalesID	1329056
SalePrice	22000.0
MachineID	1336053
ModelID	4089
datasource	132
auctioneerID	99.0
YearMade	1984
MachineHoursCurrentMeter	NaN
UsageBand	NaN
fiModelDesc	D3B
fiBaseModel	D3
fiSecondaryDesc	B
fiModelSeries	NaN
fiModelDescriptor	NaN
ProductSize	NaN
fiProductClassDesc	Track Type Tractor, Dozer - 20.0 to 75.0 Horse…
state	Florida

ProductGroup	TTT
ProductGroupDesc	Track Type Tractors
Drive_System	NaN
Enclosure	OROPS
Forks	NaN
Pad_Type	NaN
Ride_Control	NaN
Stick	NaN
Transmission	Standard
Turbocharged	NaN
Blade_Extension	NaN
Blade_Width	NaN
Enclosure_Type	NaN
Engine_Horsepower	NaN
Hydraulics	2 Valve
Pushblock	NaN
Ripper	None or Unspecified
Scarifier	NaN
Tip_Control	NaN
Tire_Size	NaN
Coupler	NaN
Coupler_System	NaN
Grouser_Tracks	NaN
Hydraulics_Flow	NaN
Track_Type	NaN
Undercarriage_Pad_Width	NaN
Stick_Length	NaN
Thumb	NaN
Pattern_Changer	NaN
Grouser_Type	NaN
Backhoe_Mounting	None or Unspecified
Blade_Type	PAT
Travel_Controls	Lever
Differential_Type	NaN
Steering_Controls	NaN
saleYear	1989
saleMonth	1
saleDay	31
saleDayOfWeek	1
saleDayOfYear	31

1.5.2 Fill missing values

First fill numeric missing values

[35]:
```
# Checking for numeric columns

for label, content in df_tmp.items():
```

```
    if pd.api.types.is_numeric_dtype(content):
        print(label)
```

```
SalesID
SalePrice
MachineID
ModelID
datasource
auctioneerID
YearMade
MachineHoursCurrentMeter
saleYear
saleMonth
saleDay
saleDayOfWeek
saleDayOfYear
```

[36]:
```
# Checking numeric columns that contain null values

for label, content in df_tmp.items():
    if pd.api.types.is_numeric_dtype(content):
        if pd.isnull(content).sum():
            print(label)
```

```
auctioneerID
MachineHoursCurrentMeter
```

[37]:
```
# Fill missing values with median
for label, content in df_tmp.items():
    if pd.api.types.is_numeric_dtype(content):
        if pd.isnull(content).sum():
            # Adding a boolean column to store whether the data contained␣
 ↪missing value or not
            df_tmp[label+"_is_missing"] = pd.isnull(content)
            # Changing to missing rows values to median values
            df_tmp[label] = content.fillna(content.median())
```

[38]:
```
# Let's check again if there is any more missing numeric values
for label, content in df_tmp.items():
    if pd.api.types.is_numeric_dtype(content):
        if pd.isnull(content).sum():
            print(label)
```

Now let's check how many missing values did the numeric columns contained

[39]:
```
df_tmp["auctioneerID_is_missing"].value_counts()
```

[39]:
```
False    392562
True      20136
Name: auctioneerID_is_missing, dtype: int64
```

[40]:
```
df_tmp["MachineHoursCurrentMeter_is_missing"].value_counts()
```

[40]:
```
True     265194
False    147504
Name: MachineHoursCurrentMeter_is_missing, dtype: int64
```

1.5.3 Filling and tuning categorical values into numbers

[41]:
```
# Check for columns which aren't numeric
for label, content in df_tmp.items():
    if not pd.api.types.is_numeric_dtype(content):
        print(label)
```

```
UsageBand
fiModelDesc
fiBaseModel
fiSecondaryDesc
fiModelSeries
fiModelDescriptor
ProductSize
fiProductClassDesc
state
ProductGroup
ProductGroupDesc
Drive_System
Enclosure
Forks
Pad_Type
Ride_Control
Stick
Transmission
Turbocharged
Blade_Extension
Blade_Width
Enclosure_Type
Engine_Horsepower
Hydraulics
Pushblock
Ripper
Scarifier
Tip_Control
Tire_Size
Coupler
Coupler_System
```

```
Grouser_Tracks
Hydraulics_Flow
Track_Type
Undercarriage_Pad_Width
Stick_Length
Thumb
Pattern_Changer
Grouser_Type
Backhoe_Mounting
Blade_Type
Travel_Controls
Differential_Type
Steering_Controls
```

[45]:
```
# Turn categorical values into numbers and fill missing
for label, content in df_tmp.items():
    if not pd.api.types.is_numeric_dtype(content):
        # Adding boolean column to store whether the row contained missing␣
↪value or not
        df_tmp[label+"_is_missing"] = pd.isnull(content)
        # Turn categorical values into numbers and +1, because the missing␣
↪value is given -1
        df_tmp[label] = pd.Categorical(content).codes +1
```

[44]:
```
pd.Categorical(df_tmp["state"]).codes
```

[44]:
```
array([43,  8,  8, …,  4,  4,  4], dtype=int8)
```

[46]:
```
df_tmp.head()
```

[46]:

	SalesID	SalePrice	MachineID	ModelID	datasource	auctioneerID	YearMade	\
0	1646770	9500.0	1126363	8434	132	18.0	1974	
1	1821514	14000.0	1194089	10150	132	99.0	1980	
2	1505138	50000.0	1473654	4139	132	99.0	1978	
3	1671174	16000.0	1327630	8591	132	99.0	1980	
4	1329056	22000.0	1336053	4089	132	99.0	1984	

	MachineHoursCurrentMeter	UsageBand	fiModelDesc	…	\
0	0.0	0	4593	…	
1	0.0	0	1820	…	
2	0.0	0	2348	…	
3	0.0	0	1819	…	
4	0.0	0	2119	…	

	Undercarriage_Pad_Width_is_missing	Stick_Length_is_missing	\
0	True	True	
1	True	True	
2	True	True	

```
3                                 True                      True
4                                 True                      True

   Thumb_is_missing  Pattern_Changer_is_missing  Grouser_Type_is_missing  \
0              True                        True                     True
1              True                        True                     True
2              True                        True                     True
3              True                        True                     True
4              True                        True                     True

   Backhoe_Mounting_is_missing  Blade_Type_is_missing  \
0                        False                  False
1                         True                   True
2                        False                  False
3                         True                   True
4                        False                  False

   Travel_Controls_is_missing  Differential_Type_is_missing  \
0                       False                          True
1                        True                         False
2                       False                          True
3                        True                         False
4                       False                          True

   Steering_Controls_is_missing
0                          True
1                         False
2                          True
3                         False
4                          True

[5 rows x 103 columns]
```

[47]:
```
df_tmp.isna().sum()
```

[47]:
```
SalesID                         0
SalePrice                       0
MachineID                       0
ModelID                         0
datasource                      0
                               ..
Backhoe_Mounting_is_missing     0
Blade_Type_is_missing           0
Travel_Controls_is_missing      0
Differential_Type_is_missing    0
Steering_Controls_is_missing    0
Length: 103, dtype: int64
```

Now that all of our data is numeric and our dataframe has no missing values, we should able to fit a machine learning model to our data

[49]:
```
%%time
from sklearn.ensemble import RandomForestRegressor
# Instantiate model

model = RandomForestRegressor(n_jobs=-1,
                              random_state=42)

# Fit the model
model.fit(df_tmp.drop("SalePrice", axis=1), df_tmp["SalePrice"])
```

```
Wall time: 3min 57s
```

[49]:
```
RandomForestRegressor(n_jobs=-1, random_state=42)
```

[50]:
```
# Score the model
model.score(df_tmp.drop("SalePrice", axis=1), df_tmp["SalePrice"])
```

[50]:
```
0.9875468079970562
```

1.6 Split the data into train and validation sets

[51]:
```
df_tmp["saleYear"]
```

[51]:
```
0         1989
1         1989
2         1989
3         1989
4         1989
          …
412693    2012
412694    2012
412695    2012
412696    2012
412697    2012
Name: saleYear, Length: 412698, dtype: int64
```

[53]:
```
df_tmp["saleYear"].value_counts()
```

[53]:
```
2009    43849
2008    39767
2011    35197
2010    33390
2007    32208
2006    21685
2005    20463
```

```
2004    19879
2001    17594
2000    17415
2002    17246
2003    15254
1998    13046
1999    12793
2012    11573
1997     9785
1996     8829
1995     8530
1994     7929
1993     6303
1992     5519
1991     5109
1989     4806
1990     4529
Name: saleYear, dtype: int64
```

[54]:
```
# Splitting data into training and validation sets
df_train = df_tmp[df_tmp["saleYear"] != 2012]
df_val = df_tmp[df_tmp["saleYear"] == 2012]

len(df_train), len(df_val)
```

[54]:
```
(401125, 11573)
```

[55]:
```
# Splitting data into X and y

X_train, y_train = df_train.drop("SalePrice", axis=1), df_train["SalePrice"]

X_valid, y_valid = df_val.drop("SalePrice", axis=1), df_val["SalePrice"]

X_train.shape, X_valid.shape, y_train.shape, y_valid.shape
```

[55]:
```
((401125, 102), (11573, 102), (401125,), (11573,))
```

[56]:
```
y_train
```

[56]:
```
0          9500.0
1         14000.0
2         50000.0
3         16000.0
4         22000.0
           ...
401120    29000.0
401121    11000.0
```

```
401122    11000.0
401123    18000.0
401124    13500.0
Name: SalePrice, Length: 401125, dtype: float64
```

1.6.1 Building a custom evaluation metric

[57]:
```
# Creating an evaluation function (the kaggle competition requires RMSLE)

from sklearn.metrics import mean_squared_log_error, mean_absolute_error,␣
↪r2_score

def rmsle(y_test, y_preds):
    """
    Calculates root mean squared log error between true and
    predicted labels
    """

    return np.sqrt(mean_squared_log_error(y_test, y_preds))

def show_scores(model):
    train_preds = model.predict(X_train)
    val_preds = model.predict(X_valid)

    scores = {
        "Train MAE": mean_absolute_error(y_train, train_preds),
        "Validation MAE": mean_absolute_error(y_valid, val_preds),
        "Train RMSLE": rmsle(y_train, train_preds),
        "Validation RMSLE": rmsle(y_valid, val_preds),
        "Train R2": r2_score(y_train, train_preds),
        "Validation R2": r2_score(y_valid, val_preds)
    }

    return scores
```

1.6.2 Testing our model on a subset (to tune Hyperparameters)

[58]:
```
%%time
model = RandomForestRegressor(n_jobs=-1,
                              random_state=42)

model.fit(X_train, y_train)
```

```
Wall time: 8min 32s
```

[58]:
```
RandomForestRegressor(n_jobs=-1, random_state=42)
```

As we can see, going through all the samples take a lot of time for the model to find patterns. One

thing we can do is to use the `max_samples` parameter to take a subset of our training data to fit the model.

[59]:
```
# Chainging max samples value
model = RandomForestRegressor(n_jobs=-1,
                              random_state=42,
                              max_samples=10000)
```

[60]:
```
%%time
# Now we will check how cutting down the amount of samples to be trained␣
 ↪reduces time for the model to fit
model.fit(X_train, y_train)
```

```
Wall time: 9.37 s
```

[60]:
```
RandomForestRegressor(max_samples=10000, n_jobs=-1, random_state=42)
```

[61]:
```
show_scores(model)
```

[61]:
```
{'Train MAE': 5561.2988092240585,
 'Validation MAE': 7177.26365505919,
 'Train RMSLE': 0.257745378256977,
 'Validation RMSLE': 0.29362638671089003,
 'Train R2': 0.8606658995199189,
 'Validation R2': 0.8320374995090507}
```

1.6.3 Hyperparameter tuning with RandomizedSearchCV

[62]:
```
%%time
from sklearn.model_selection import RandomizedSearchCV

# Different random forest hyperparameters
rf_grid = {
    "n_estimators": np.arange(10,100,10),
    "max_samples": [10000],
    "max_depth": [None, 3, 5, 10],
    "min_samples_leaf": np.arange(1,20,2),
    "min_samples_split": np.arange(2,20,2),
    "max_features": [0.5, 1, "sqrt", "auto"]
}

# Instantiate a randomized search model

rs_model = RandomizedSearchCV(RandomForestRegressor(n_jobs=-1,
                                                    random_state=42),
                              param_distributions=rf_grid,
                              n_iter=100,
                              cv=5,
```

```
                                verbose=2)

# Fit the data into randomized search model
rs_model.fit(X_train, y_train)
```

```
Fitting 5 folds for each of 100 candidates, totalling 500 fits
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=80; total time=  24.2s
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=80; total time=   7.3s
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=80; total time=   6.7s
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=80; total time=   6.6s
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=80; total time=   6.6s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=10, n_estimators=90; total time=   2.7s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=10, n_estimators=90; total time=   2.8s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=10, n_estimators=90; total time=   2.8s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=10, n_estimators=90; total time=   2.7s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=10, n_estimators=90; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=40; total time=   2.6s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=40; total time=   2.6s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=40; total time=   2.8s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=12, n_estimators=40; total time=   2.6s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=14, n_estimators=20; total time=   2.3s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=14, n_estimators=20; total time=   2.5s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=14, n_estimators=20; total time=   2.3s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=14, n_estimators=20; total time=   2.6s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=14, n_estimators=20; total time=   2.6s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=6, n_estimators=60; total time=   3.1s
```

```
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=6, n_estimators=60; total time=   3.1s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=6, n_estimators=60; total time=   2.8s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=6, n_estimators=60; total time=   2.8s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=6, n_estimators=60; total time=   2.9s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=20; total time=   2.3s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=20; total time=   2.3s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=20; total time=   2.3s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=20; total time=   2.5s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=20; total time=   2.6s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=40; total time=   2.4s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=40; total time=   2.2s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=40; total time=   2.2s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=40; total time=   2.2s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=40; total time=   2.4s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=8, n_estimators=40; total time=   2.9s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=8, n_estimators=40; total time=   2.8s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=8, n_estimators=40; total time=   2.6s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=8, n_estimators=40; total time=   2.6s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=8, n_estimators=40; total time=   2.6s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=10; total time=   2.0s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=10; total time=   1.9s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=10; total time=   2.0s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=10; total time=   2.0s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=10; total time=   2.0s
```

```
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=90; total time=   2.4s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=90; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=90; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=90; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=90; total time=   2.5s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=2, n_estimators=20; total time=   2.3s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=2, n_estimators=20; total time=   2.2s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=2, n_estimators=20; total time=   2.2s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=2, n_estimators=20; total time=   2.3s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=2, n_estimators=20; total time=   2.2s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=12, n_estimators=90; total time=   3.6s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=12, n_estimators=90; total time=   3.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=12, n_estimators=90; total time=   3.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=12, n_estimators=90; total time=   3.6s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=12, n_estimators=90; total time=   3.5s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=70; total time=   6.7s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=70; total time=   6.6s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=70; total time=   6.7s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=70; total time=   6.4s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=70; total time=   6.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=19, min_samples_split=4, n_estimators=60; total time=   2.7s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=19, min_samples_split=4, n_estimators=60; total time=   3.0s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=19, min_samples_split=4, n_estimators=60; total time=   2.8s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=19, min_samples_split=4, n_estimators=60; total time=   2.9s
```

```
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=19, min_samples_split=4, n_estimators=60; total time=   2.8s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.1s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.1s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.3s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.1s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=20; total time=   3.2s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=20; total time=   3.1s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=20; total time=   3.2s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=20; total time=   3.2s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=11, min_samples_split=16, n_estimators=20; total time=   3.4s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=40; total time=   2.5s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=40; total time=   2.6s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=40; total time=   2.5s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=90; total time=   3.6s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=90; total time=   3.7s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=90; total time=   3.6s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=90; total time=   3.6s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=90; total time=   3.5s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=30; total time=   2.2s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=30; total time=   2.3s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=30; total time=   2.3s
```

```
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=30; total time=   2.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=30; total time=   2.1s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=20; total time=   2.1s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=20; total time=   2.1s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=20; total time=   2.0s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=20; total time=   2.1s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=20; total time=   2.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=70; total time=   2.4s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=70; total time=   2.5s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=70; total time=   2.5s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=70; total time=   2.4s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=70; total time=   2.3s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=40; total time=   2.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=40; total time=   2.5s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=40; total time=   2.4s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=40; total time=   2.1s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=40; total time=   2.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=40; total time=   2.1s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=40; total time=   2.3s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=12, n_estimators=50; total time=   2.3s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=12, n_estimators=50; total time=   2.3s
```

```
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=12, n_estimators=50; total time=   2.4s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=12, n_estimators=50; total time=   2.2s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=12, n_estimators=50; total time=   2.3s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=3,
min_samples_split=8, n_estimators=60; total time=   2.2s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=3,
min_samples_split=8, n_estimators=60; total time=   2.4s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=3,
min_samples_split=8, n_estimators=60; total time=   2.5s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=3,
min_samples_split=8, n_estimators=60; total time=   2.3s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=3,
min_samples_split=8, n_estimators=60; total time=   2.2s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=6, n_estimators=20; total time=   3.0s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=6, n_estimators=20; total time=   3.2s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=6, n_estimators=20; total time=   3.1s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=6, n_estimators=20; total time=   3.2s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=6, n_estimators=20; total time=   3.1s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=10, n_estimators=60; total time=   2.4s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=10, n_estimators=60; total time=   2.1s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=10, n_estimators=60; total time=   2.1s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=10, n_estimators=60; total time=   2.0s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=10, n_estimators=60; total time=   1.9s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=14, n_estimators=40; total time=   2.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=14, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=14, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=14, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=14, n_estimators=40; total time=   2.7s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=60; total time=   2.1s
```

```
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=60; total time=   2.1s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=60; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=60; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=60; total time=   2.2s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=17, min_samples_split=6, n_estimators=90; total time=   3.2s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=17, min_samples_split=6, n_estimators=90; total time=   3.2s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=17, min_samples_split=6, n_estimators=90; total time=   3.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=17, min_samples_split=6, n_estimators=90; total time=   3.3s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=17, min_samples_split=6, n_estimators=90; total time=   3.4s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=20; total time=   2.1s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=20; total time=   2.1s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=20; total time=   2.1s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=20; total time=   2.0s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=20; total time=   2.2s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=14, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=40; total time=   3.7s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=40; total time=   3.5s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=40; total time=   3.5s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=40; total time=   3.5s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=40; total time=   3.6s
```

```
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=14, n_estimators=30; total time=   2.9s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=14, n_estimators=30; total time=   2.8s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=14, n_estimators=30; total time=   3.0s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=14, n_estimators=30; total time=   2.8s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=14, n_estimators=30; total time=   2.8s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=16, n_estimators=10; total time=   2.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=16, n_estimators=10; total time=   2.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=16, n_estimators=10; total time=   2.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=16, n_estimators=10; total time=   2.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=7,
min_samples_split=16, n_estimators=10; total time=   2.0s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=10; total time=   2.1s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=10; total time=   2.0s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=10; total time=   2.1s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=10; total time=   2.1s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=10; total time=   2.1s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=10; total time=   2.2s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=10; total time=   2.3s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=10; total time=   2.3s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=10; total time=   2.4s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=15, min_samples_split=2, n_estimators=10; total time=   2.1s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=80; total time=   2.8s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=80; total time=   3.1s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=80; total time=   2.9s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=80; total time=   2.8s
```

```
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=80; total time=   2.8s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=14, n_estimators=30; total time=   2.8s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=14, n_estimators=30; total time=   2.6s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=14, n_estimators=30; total time=   2.6s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=14, n_estimators=30; total time=   2.6s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=17,
min_samples_split=14, n_estimators=30; total time=   2.6s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=80; total time=   3.4s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=80; total time=   3.6s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=80; total time=   3.6s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=80; total time=   3.5s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=2, n_estimators=80; total time=   3.6s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=7, min_samples_split=4, n_estimators=30; total time=   3.1s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=7, min_samples_split=4, n_estimators=30; total time=   2.9s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=7, min_samples_split=4, n_estimators=30; total time=   3.3s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=7, min_samples_split=4, n_estimators=30; total time=   3.1s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=7, min_samples_split=4, n_estimators=30; total time=   3.3s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=40; total time=   2.4s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=40; total time=   2.2s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=40; total time=   2.2s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=40; total time=   2.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=40; total time=   2.0s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=70; total time=   3.4s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=70; total time=   3.4s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=70; total time=   3.5s
```

```
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=70; total time=   3.3s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=70; total time=   3.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=10, n_estimators=30; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=10, n_estimators=30; total time=   2.0s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=10, n_estimators=30; total time=   2.0s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=10, n_estimators=30; total time=   2.0s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=10, n_estimators=30; total time=   2.0s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=8, n_estimators=70; total time=   4.0s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=8, n_estimators=70; total time=   4.6s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=8, n_estimators=70; total time=   4.0s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=8, n_estimators=70; total time=   4.0s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=8, n_estimators=70; total time=   4.0s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=30; total time=   2.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=30; total time=   2.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=30; total time=   2.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=30; total time=   2.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=30; total time=   2.2s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=1,
min_samples_split=6, n_estimators=30; total time=   2.7s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=1,
min_samples_split=6, n_estimators=30; total time=   2.6s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=1,
min_samples_split=6, n_estimators=30; total time=   2.8s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=1,
min_samples_split=6, n_estimators=30; total time=   2.7s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=1,
min_samples_split=6, n_estimators=30; total time=   2.6s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=18, n_estimators=70; total time=   3.2s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=18, n_estimators=70; total time=   3.4s
```

```
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=18, n_estimators=70; total time=   3.3s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=18, n_estimators=70; total time=   3.3s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=18, n_estimators=70; total time=   3.5s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=80; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=80; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=80; total time=   2.3s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=80; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=17,
min_samples_split=2, n_estimators=80; total time=   2.2s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=6, n_estimators=30; total time=   3.0s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=6, n_estimators=30; total time=   3.1s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=6, n_estimators=30; total time=   3.1s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=6, n_estimators=30; total time=   3.1s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=6, n_estimators=30; total time=   3.1s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=8, n_estimators=50; total time=   3.5s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=8, n_estimators=50; total time=   3.5s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=8, n_estimators=50; total time=   3.5s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=8, n_estimators=50; total time=   3.4s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=8, n_estimators=50; total time=   3.6s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=50; total time=   2.6s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=50; total time=   2.4s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=50; total time=   2.5s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=50; total time=   2.4s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=13,
min_samples_split=4, n_estimators=50; total time=   2.4s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=90; total time=   2.5s
```

```
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=90; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=90; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=90; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=2, n_estimators=90; total time=   2.7s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=17, min_samples_split=18, n_estimators=60; total time=   4.1s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=17, min_samples_split=18, n_estimators=60; total time=   5.3s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=17, min_samples_split=18, n_estimators=60; total time=   4.1s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=17, min_samples_split=18, n_estimators=60; total time=   4.1s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=17, min_samples_split=18, n_estimators=60; total time=   4.1s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=9,
min_samples_split=10, n_estimators=30; total time=   2.3s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=9,
min_samples_split=10, n_estimators=30; total time=   2.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=9,
min_samples_split=10, n_estimators=30; total time=   2.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=9,
min_samples_split=10, n_estimators=30; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=9,
min_samples_split=10, n_estimators=30; total time=   2.5s
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=18, n_estimators=70; total time=   6.1s
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=18, n_estimators=70; total time=   6.0s
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=18, n_estimators=70; total time=   6.0s
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=18, n_estimators=70; total time=   6.0s
[CV] END max_depth=10, max_features=auto, max_samples=10000,
min_samples_leaf=19, min_samples_split=18, n_estimators=70; total time=   5.9s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=18, n_estimators=10; total time=   2.0s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=18, n_estimators=10; total time=   2.0s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=18, n_estimators=10; total time=   2.0s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=18, n_estimators=10; total time=   2.0s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=18, n_estimators=10; total time=   2.0s
```

```
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=12, n_estimators=10; total time=   1.9s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=12, n_estimators=10; total time=   2.1s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=12, n_estimators=10; total time=   2.2s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=12, n_estimators=10; total time=   2.1s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=12, n_estimators=10; total time=   2.2s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=70; total time=   3.7s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=70; total time=   3.6s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=70; total time=   3.7s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=70; total time=   3.9s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=70; total time=   3.7s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=6, n_estimators=20; total time=   2.3s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=6, n_estimators=20; total time=   2.4s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=6, n_estimators=20; total time=   2.3s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=6, n_estimators=20; total time=   2.3s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=6, n_estimators=20; total time=   2.3s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=60; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=60; total time=   2.1s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=60; total time=   1.9s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=60; total time=   2.1s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=18, n_estimators=60; total time=   2.0s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=5, min_samples_split=14, n_estimators=10; total time=   2.5s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=5, min_samples_split=14, n_estimators=10; total time=   2.5s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=5, min_samples_split=14, n_estimators=10; total time=   2.7s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=5, min_samples_split=14, n_estimators=10; total time=   2.5s
```

```
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=5, min_samples_split=14, n_estimators=10; total time=   2.7s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=70; total time=   3.9s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=70; total time=   4.2s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=70; total time=   4.0s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=70; total time=   4.2s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=70; total time=   4.0s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=80; total time=   2.7s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=80; total time=   2.9s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=80; total time=   2.7s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=80; total time=   2.6s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=80; total time=   2.7s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=50; total time=   2.4s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=50; total time=   2.4s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=50; total time=   2.6s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=50; total time=   2.5s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=16, n_estimators=50; total time=   2.4s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=80; total time=   3.6s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=80; total time=   3.7s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=80; total time=   3.7s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=80; total time=   3.9s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=18, n_estimators=80; total time=   4.3s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=40; total time=   2.9s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=40; total time=   2.9s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=40; total time=   2.3s
```

```
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=16, n_estimators=40; total time=   2.5s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=50; total time=   2.6s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=50; total time=   2.9s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=50; total time=   2.5s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=50; total time=   3.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=16, n_estimators=50; total time=   2.3s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=20; total time=   3.0s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=20; total time=   3.3s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=20; total time=   3.3s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=20; total time=   3.3s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=20; total time=   3.2s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=10; total time=   2.2s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=10; total time=   2.2s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=10; total time=   2.1s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=10; total time=   2.3s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=10; total time=   2.3s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=14, n_estimators=80; total time=   5.4s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=14, n_estimators=80; total time=   5.0s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=14, n_estimators=80; total time=   5.3s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=14, n_estimators=80; total time=   5.4s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=11,
min_samples_split=14, n_estimators=80; total time=   5.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=14, n_estimators=20; total time=   2.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=14, n_estimators=20; total time=   2.1s
```

```
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=14, n_estimators=20; total time=   2.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=14, n_estimators=20; total time=   2.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=3,
min_samples_split=14, n_estimators=20; total time=   2.2s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=12, n_estimators=70; total time=   5.0s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=12, n_estimators=70; total time=   4.9s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=12, n_estimators=70; total time=   5.0s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=12, n_estimators=70; total time=   4.9s
[CV] END max_depth=5, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=12, n_estimators=70; total time=   5.0s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=18, n_estimators=80; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=18, n_estimators=80; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=18, n_estimators=80; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=18, n_estimators=80; total time=   2.2s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=18, n_estimators=80; total time=   2.3s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=10; total time=   2.3s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=10; total time=   2.4s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=10; total time=   2.3s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=10; total time=   2.3s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=15,
min_samples_split=16, n_estimators=10; total time=   2.3s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=4, n_estimators=60; total time=   2.5s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=4, n_estimators=60; total time=   2.4s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=4, n_estimators=60; total time=   2.4s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=4, n_estimators=60; total time=   2.4s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=4, n_estimators=60; total time=   2.4s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=90; total time=   2.3s
```

```
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=90; total time=   2.6s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=90; total time=   2.6s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=90; total time=   2.5s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=1,
min_samples_split=12, n_estimators=90; total time=   2.6s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=90; total time=   2.4s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=90; total time=   2.3s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=90; total time=   2.3s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=90; total time=   2.5s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=90; total time=   2.6s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=18, n_estimators=20; total time=   3.2s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=18, n_estimators=20; total time=   3.3s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=18, n_estimators=20; total time=   3.2s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=18, n_estimators=20; total time=   3.2s
[CV] END max_depth=None, max_features=auto, max_samples=10000,
min_samples_leaf=13, min_samples_split=18, n_estimators=20; total time=   3.4s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=50; total time=   2.3s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=50; total time=   2.6s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=50; total time=   2.5s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=50; total time=   2.4s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=19,
min_samples_split=16, n_estimators=50; total time=   2.2s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=14, n_estimators=90; total time=   4.5s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=14, n_estimators=90; total time=   4.5s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=14, n_estimators=90; total time=   4.5s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=14, n_estimators=90; total time=   4.5s
[CV] END max_depth=5, max_features=0.5, max_samples=10000, min_samples_leaf=15,
min_samples_split=14, n_estimators=90; total time=   4.5s
```

```
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=50; total time=   4.0s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=50; total time=   4.0s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=50; total time=   3.9s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=50; total time=   4.3s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=12, n_estimators=50; total time=   4.2s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=90; total time=   5.1s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=90; total time=   4.8s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=90; total time=   4.9s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=90; total time=   5.0s
[CV] END max_depth=3, max_features=auto, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=90; total time=   4.9s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=20; total time=   2.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=20; total time=   2.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=20; total time=   2.1s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=20; total time=   2.0s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=15,
min_samples_split=6, n_estimators=20; total time=   2.1s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=60; total time=   2.7s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=60; total time=   2.6s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=60; total time=   3.0s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=60; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=11,
min_samples_split=6, n_estimators=60; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=16, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=16, n_estimators=40; total time=   2.5s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=16, n_estimators=40; total time=   2.4s
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=16, n_estimators=40; total time=   2.3s
```

```
[CV] END max_depth=10, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=16, n_estimators=40; total time=   2.3s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=8, n_estimators=60; total time=   3.3s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=8, n_estimators=60; total time=   3.0s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=8, n_estimators=60; total time=   3.2s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=8, n_estimators=60; total time=   2.8s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=13, min_samples_split=8, n_estimators=60; total time=   2.9s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=10, n_estimators=40; total time=   2.8s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=10, n_estimators=40; total time=   2.8s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=10, n_estimators=40; total time=   2.9s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=10, n_estimators=40; total time=   2.7s
[CV] END max_depth=None, max_features=sqrt, max_samples=10000,
min_samples_leaf=3, min_samples_split=10, n_estimators=40; total time=   2.6s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=60; total time=   5.8s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=60; total time=   5.8s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=60; total time=   5.7s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=60; total time=   6.2s
[CV] END max_depth=10, max_features=auto, max_samples=10000, min_samples_leaf=3,
min_samples_split=16, n_estimators=60; total time=   5.9s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=17,
min_samples_split=10, n_estimators=40; total time=   2.6s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=17,
min_samples_split=10, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=17,
min_samples_split=10, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=17,
min_samples_split=10, n_estimators=40; total time=   2.3s
[CV] END max_depth=5, max_features=sqrt, max_samples=10000, min_samples_leaf=17,
min_samples_split=10, n_estimators=40; total time=   2.6s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=5,
min_samples_split=2, n_estimators=90; total time=   4.1s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=5,
min_samples_split=2, n_estimators=90; total time=   3.9s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=5,
min_samples_split=2, n_estimators=90; total time=   4.0s
```

```
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=5,
min_samples_split=2, n_estimators=90; total time=   3.9s
[CV] END max_depth=3, max_features=0.5, max_samples=10000, min_samples_leaf=5,
min_samples_split=2, n_estimators=90; total time=   3.8s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=50; total time=   3.9s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=50; total time=   4.0s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=50; total time=   4.0s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=50; total time=   4.0s
[CV] END max_depth=10, max_features=0.5, max_samples=10000, min_samples_leaf=7,
min_samples_split=4, n_estimators=50; total time=   4.1s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=30; total time=   2.3s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=30; total time=   2.3s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=30; total time=   2.1s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=30; total time=   2.1s
[CV] END max_depth=5, max_features=1, max_samples=10000, min_samples_leaf=5,
min_samples_split=4, n_estimators=30; total time=   2.3s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=30; total time=   2.2s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=30; total time=   2.2s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=30; total time=   2.2s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=30; total time=   2.5s
[CV] END max_depth=None, max_features=1, max_samples=10000, min_samples_leaf=9,
min_samples_split=4, n_estimators=30; total time=   2.5s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=4, n_estimators=80; total time=   3.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=4, n_estimators=80; total time=   3.2s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=4, n_estimators=80; total time=   3.6s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=4, n_estimators=80; total time=   3.9s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000,
min_samples_leaf=11, min_samples_split=4, n_estimators=80; total time=   3.5s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=5, min_samples_split=8, n_estimators=20; total time=   2.9s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=5, min_samples_split=8, n_estimators=20; total time=   2.8s
```

```
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=5, min_samples_split=8, n_estimators=20; total time=   3.0s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=5, min_samples_split=8, n_estimators=20; total time=   2.9s
[CV] END max_depth=None, max_features=0.5, max_samples=10000,
min_samples_leaf=5, min_samples_split=8, n_estimators=20; total time=   3.1s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=2, n_estimators=60; total time=   3.1s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=2, n_estimators=60; total time=   3.1s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=2, n_estimators=60; total time=   3.0s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=2, n_estimators=60; total time=   3.1s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=7,
min_samples_split=2, n_estimators=60; total time=   3.1s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=10, n_estimators=80; total time=   2.5s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=10, n_estimators=80; total time=   2.4s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=10, n_estimators=80; total time=   2.7s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=10, n_estimators=80; total time=   2.5s
[CV] END max_depth=3, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=10, n_estimators=80; total time=   2.6s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=8, n_estimators=20; total time=   2.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=8, n_estimators=20; total time=   2.4s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=8, n_estimators=20; total time=   2.2s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=8, n_estimators=20; total time=   2.3s
[CV] END max_depth=10, max_features=sqrt, max_samples=10000, min_samples_leaf=5,
min_samples_split=8, n_estimators=20; total time=   2.3s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=90; total time=   2.4s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=90; total time=   2.5s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=90; total time=   2.5s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=90; total time=   2.3s
[CV] END max_depth=3, max_features=1, max_samples=10000, min_samples_leaf=13,
min_samples_split=12, n_estimators=90; total time=   2.3s
Wall time: 25min 41s
```

[62]:
```
RandomizedSearchCV(cv=5,
                   estimator=RandomForestRegressor(n_jobs=-1, random_state=42),
                   n_iter=100,
                   param_distributions={'max_depth': [None, 3, 5, 10],
                                        'max_features': [0.5, 1, 'sqrt',
                                                         'auto'],
                                        'max_samples': [10000],
                                        'min_samples_leaf': array([ 1,  3,  5,
7,  9, 11, 13, 15, 17, 19]),
                                        'min_samples_split': array([ 2,  4,  6,
8, 10, 12, 14, 16, 18]),
                                        'n_estimators': array([10, 20, 30, 40,
50, 60, 70, 80, 90])},
                   verbose=2)
```

[63]:
```
# Let's check the best hyperparameters of our model
rs_model.best_params_
```

[63]:
```
{'n_estimators': 20,
 'min_samples_split': 8,
 'min_samples_leaf': 5,
 'max_samples': 10000,
 'max_features': 0.5,
 'max_depth': None}
```

[65]:
```
#Evaluate the RandomizedSearch model on 10,000 samples
show_scores(rs_model)
```

[65]:
```
{'Train MAE': 6139.510732219108,
 'Validation MAE': 7496.72475593148,
 'Train RMSLE': 0.2774216427640643,
 'Validation RMSLE': 0.30466519366956274,
 'Train R2': 0.8316709696905946,
 'Validation R2': 0.8165568169833842}
```

[67]:
```
### Fitting our model with best hyperparameters with the training set
ideal_model = RandomForestRegressor(n_estimators=20,
                                    min_samples_leaf=5,
                                    min_samples_split=8,
                                    max_samples=None,
                                    max_depth=None,
                                    max_features=0.5,
                                    random_state=42)

ideal_model.fit(X_train , y_train)
```

[67]:
```
RandomForestRegressor(max_features=0.5, min_samples_leaf=5, min_samples_split=8,
                      n_estimators=20)
```

[68]:
```
# Scores for ideal model, trained on all the data
show_scores(ideal_model)
```

[68]:
```
{'Train MAE': 3204.039037495885,
 'Validation MAE': 6023.030862495653,
 'Train RMSLE': 0.15772858775401682,
 'Validation RMSLE': 0.2472564453034715,
 'Train R2': 0.9495967519802064,
 'Validation R2': 0.8772931987886761}
```

After iterating for 100 times, with the help of hyperparameter tuninh, we have a significantly improved model.

We can now fit our test data to this model to predict the bulldozer prices

[ ]:
```

```