In [5]:
```
from datetime import datetime
from scipy.stats import skew  # for some statistics
from scipy.special import boxcox1p
from scipy.stats import boxcox_normmax
from sklearn.linear_model import ElasticNetCV, LassoCV, RidgeCV
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
from sklearn.svm import SVR
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import RobustScaler
from sklearn.model_selection import KFold, cross_val_score
from sklearn.metrics import mean_squared_error
from mlxtend.regressor import StackingCVRegressor
from xgboost import XGBRegressor
from lightgbm import LGBMRegressor
import matplotlib.pyplot as plt
import scipy.stats as stats
import sklearn.linear_model as linear_model
import seaborn as sns
from sklearn.manifold import TSNE
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import tensorflow as tf
import openpyxl
from keras.models import Sequential
from keras.layers import Dense
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.compose import ColumnTransformer
from sklearn.metrics import r2_score
from sklearn.neural_network import MLPRegressor
import os

import warnings
warnings.filterwarnings('ignore')
```

In [6]:
```
dataset = pd.read_csv("analysis_new.csv")
df = pd.read_csv("testnya_new.csv")
X_train = dataset.iloc[:, :-1].values
y_train = dataset.iloc[:, -1].values
X_test = df.iloc[:, :].values
X_train.shape,y_train.shape,X_test.shape
```

Out[6]:
```
((23268, 25), (23268,), (2208, 25))
```

In [7]:
```
k_fold = KFold(n_splits = 15, random_state = 11, shuffle = True)
def cv_rmse(model, X = X_train):
    rmse = np.sqrt(-cross_val_score(model, X_train, y_train, scoring = "neg_mean_sq
    return rmse
```

```
def rmsle(y, y_pred):
    return np.sqrt(mean_squared_error(y, y_pred))
```

XGBoost

In [8]:
```
xgb = make_pipeline(RobustScaler(),
                    XGBRegressor(colsample_bytree = 0.5, n_estimators = 6000,
                                 max_depth = 4, learning_rate = 0.01, gamma = 0.45,
                                 subsample = 0.5, random_state = 11, reg_alpha = 0.
                                 reg_lambda = None, nthread = -1))
```

In [9]:
```
# get CV score of the xgb model
score = cv_rmse(xgb)
print("Xgboost model's cross validation score: ", score.mean())
```

```
Xgboost model's cross validation score:  0.12762977988402593
```

LightGBM

In [10]:
```
lgbm = make_pipeline(RobustScaler(),
                     LGBMRegressor(num_leaves = 6, bagging_fraction = 0.7,
                                   bagging_freq = 4, min_sum_hessian_in_leaf = 11,
                                   learning_rate = 0.01, n_estimators = 7500, max_b
                                   random_state = 11))
```

In [11]:
```
# get CV score of the lgbm model
score = cv_rmse(lgbm)
print("Light GBM model's cross validation score: ", score.mean())
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing wa
s 0.002939 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 2415
[LightGBM] [Info] Number of data points in the train set: 21716, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224142
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002818 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2421
[LightGBM] [Info] Number of data points in the train set: 21716, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224456
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
```

```
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing wa
s 0.001679 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 2421
[LightGBM] [Info] Number of data points in the train set: 21716, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.223153
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.004397 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2418
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224189
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
```

```
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing wa
s 0.003582 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 2420
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.223712
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.004775 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2418
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224569
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.006393 seconds.
You can set `force_col_wise=true` to remove the overhead.
```

```
[LightGBM] [Info] Total Bins 2417
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224502
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002632 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2418
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224024
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.005400 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2417
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.223709
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
```

```
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing wa
s 0.002061 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 2421
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224264
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.003811 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2417
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224783
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
```

```
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002816 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2417
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.223318
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing wa
s 0.001825 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 2416
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224119
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing wa
s 0.001908 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 2420
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224842
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.004606 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2420
[LightGBM] [Info] Number of data points in the train set: 21717, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.225066
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
Light GBM model's cross validation score:  0.10259560051214617
```

Ridge

In [12]:

```
ridge = make_pipeline(RobustScaler(),
                      RidgeCV(alphas = [1e-10, 1e-8, 1e-5, 1e-2, 9e-4,
                                        5e-4, 3e-4, 1e-4, 1e-3, 1e-
                                        0.3, 0.6, 1, 3, 5, 7, 14, 1
                                        45, 50, 70, 90], cv = k_fol
```

In [13]:
```
# get CV score of the ridge model
score = cv_rmse(ridge)
print("Ridge model's cross validation score: ", score.mean())
```

```
Ridge model's cross validation score:  0.14882459005634863
```

MLP

In [14]:
```
# Define the pipeline
mlp = make_pipeline(
    RobustScaler(),
    MLPRegressor(random_state=42,hidden_layer_sizes= (200,200,200),
                 activation='relu')
)
```

In [15]:
```
# get CV score of the MLP model
score = cv_rmse(mlp)
print("Ridge model's cross validation score: ", score.mean())
```

```
Ridge model's cross validation score:  0.08914765048906431
```

LASSO

In [16]:
```
lasso = make_pipeline(RobustScaler(),
                      LassoCV(alphas = [1e-10, 1e-8, 1e-5, 1e-2, 9e-4,
                                                        5e-4, 3e-4, 1e-4, 1e-3, 1e-
                                                        0.3, 0.6, 1, 3, 5, 7, 14, 1
                                                        45, 50, 70, 90], n_jobs = -
```

In [17]:
```
# get CV score of the lasso model
score = cv_rmse(lasso)
print("Lasso model's cross validation score: ", score.mean())
```

```
Lasso model's cross validation score:  0.14894819894521025
```

ElasticNet

In [18]:
```
elasticnet = make_pipeline(RobustScaler(),
                           ElasticNetCV(max_iter=1000,
                                        alphas=[0.0001, 0.0002, 0.0003, 0.0004, 0.0
                                        cv=k_fold,
                                        l1_ratio=[0.8, 0.85, 0.9, 0.95, 0.99, 1]))
```

In [19]:
```
# get CV score of the lasso model
score = cv_rmse(elasticnet)
print("Elastic Net model's cross validation score: ", score.mean())
```

```
Elastic Net model's cross validation score:  0.14894766070155158
```

Support Vector Regression

In [20]:
```
svr = make_pipeline(RobustScaler(), SVR(C = 30, gamma = 0.0002, epsilon = 0.009))
```

In [21]:
```
# get CV score of the svr model
score = cv_rmse(svr)
print("Support vector machines model's cross validation score: ", score.mean())
```

```
Support vector machines model's cross validation score:  0.14357722807506043
```

Gradient Boosting

In [22]:
```
gbr = make_pipeline(RobustScaler(),
                    GradientBoostingRegressor(n_estimators = 700, learning_rate = 0
                                              max_depth = 5, min_samples_split = 12
                                              loss = "huber", max_features = "sqrt"
```

In [23]:
```
# get CV score of the gbr model
score = cv_rmse(gbr)
print("Gradient boosting model's cross validation score: ", score.mean())
```

```
Gradient boosting model's cross validation score:  0.123617553873992
```

Random Forest

In [24]:
```
rf = make_pipeline(RobustScaler(),
                   RandomForestRegressor(n_estimators = 250, max_depth = 15,
                                         min_samples_split = 6, min_samples_leaf =
                                         random_state = 11))
```

In [25]:
```
# get CV score of the rf model
score = cv_rmse(rf)
print("Random forest model's cross validation score: ", score.mean())
```

```
Random forest model's cross validation score:  0.11054004733472847
```

Stacked Model

In [27]:
```
stacked = StackingCVRegressor(regressors = (xgb, lgbm, ridge, svr, lasso, elasticne
                              meta_regressor = xgb, use_features_in_secondary = Tru

stackedv2 = StackingCVRegressor(regressors = (xgb, lgbm, ridge, svr, lasso, elastic
                              meta_regressor = mlp, use_features_in_secondary = Tru

stackedv3 = StackingCVRegressor(regressors = (xgb, lgbm, ridge, svr, lasso, elastic
                              meta_regressor = lgbm, use_features_in_secondary = Tr
```

Fit all model

In [28]:
```
mlp_model = mlp.fit(X_train, y_train)

#RMSLE score of the gbr model on full train data
mlp_score = rmsle(y_train, mlp_model.predict(X_train))
print("RMSLE score of MLP model on full data:", mlp_score)
```

```
RMSLE score of MLP model on full data: 0.07067490704753682
```

In [29]:
```
gbr_model = gbr.fit(X_train, y_train)

#RMSLE score of the gbr model on full train data
gbr_score = rmsle(y_train, gbr_model.predict(X_train))
print("RMSLE score of xgboost model on full data:", gbr_score)
```

```
RMSLE score of xgboost model on full data: 0.11789715815680625
```

In [30]:
```
rf_model = rf.fit(X_train, y_train)

#RMSLE score of the rf model on full train data
rf_score = rmsle(y_train, rf_model.predict(X_train))
print("RMSLE score of random forest model on full data:", rf_score)
```

```
RMSLE score of random forest model on full data: 0.07816864234436895
```

In [31]:
```
en_model = elasticnet.fit(X_train, y_train)

#RMSLE score of the gbr model on full train data
en_score = rmsle(y_train, en_model.predict(X_train))
print("RMSLE score of xgboost model on full data:", en_score)
```

```
RMSLE score of xgboost model on full data: 0.1488297724050217
```

In [32]:
```
svr_model = svr.fit(X_train, y_train)

#RMSLE score of the svr model on full train data
svr_score = rmsle(y_train, svr_model.predict(X_train))
print("RMSLE score of svr model on full data:", svr_score)
```

```
RMSLE score of svr model on full data: 0.14312510875741458
```

In [33]:
```
xgb_model = xgb.fit(X_train, y_train)

#RMSLE score of the xgb model on full train data
xgb_score = rmsle(y_train, xgb_model.predict(X_train))
print("RMSLE score of xgboost model on full data:", xgb_score)
```

```
RMSLE score of xgboost model on full data: 0.12402484437333865
```

In [34]:
```
lgbm_model = lgbm.fit(X_train, y_train)

#RMSLE score of the lgbm model on full train data
lgbm_score = rmsle(y_train, lgbm_model.predict(X_train))
print("RMSLE score of lgbm model on full data:", lgbm_score)
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.007075 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2421
[LightGBM] [Info] Number of data points in the train set: 23268, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224190
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
RMSLE score of lgbm model on full data: 0.09241028583720148
```

In [35]:
```
ridge_model = ridge.fit(X_train, y_train)

#RMSLE score of the ridge model on full train data
ridge_score = rmsle(y_train, ridge_model.predict(X_train))
print("RMSLE score of ridge model on full data:", ridge_score)
```

```
RMSLE score of ridge model on full data: 0.14869150858310293
```

In [36]:
```
lasso_model = lasso.fit(X_train, y_train)

#RMSLE score of the lasso model on full train data
lasso_score = rmsle(y_train, lasso_model.predict(X_train))
print("RMSLE score of lasso model on full data:", lasso_score)
```

```
RMSLE score of lasso model on full data: 0.1488194827986601
```

In [37]:
```
stacked_model = stacked.fit(np.array(X_train), np.array(y_train))

#RMSLE score of the stacked model on full train data
stacked_score = rmsle(y_train, stacked_model.predict(X_train))
print("RMSLE score of stacked models on full data:", stacked_score)
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.004206 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2413
[LightGBM] [Info] Number of data points in the train set: 18614, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224817
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing wa
s 0.001226 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 2412
[LightGBM] [Info] Number of data points in the train set: 18614, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224017
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
```

```
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing wa
s 0.002172 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 2411
[LightGBM] [Info] Number of data points in the train set: 18614, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.223798
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002357 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2411
[LightGBM] [Info] Number of data points in the train set: 18615, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224152
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
```

```
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.003599 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2414
[LightGBM] [Info] Number of data points in the train set: 18615, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224165
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.003057 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2421
[LightGBM] [Info] Number of data points in the train set: 23268, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224190
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
RMSLE score of stacked models on full data: 0.06690970552471756
```

In [38]:

```
stacked_model2 = stackedv2.fit(np.array(X_train), np.array(y_train))

#RMSLE score of the stacked model on full train data
stacked_score = rmsle(y_train, stacked_model2.predict(X_train))
print("RMSLE score of stacked models on full data:", stacked_score)
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002555 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2409
[LightGBM] [Info] Number of data points in the train set: 18614, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224606
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002265 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2417
[LightGBM] [Info] Number of data points in the train set: 18614, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.225306
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002373 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2413
[LightGBM] [Info] Number of data points in the train set: 18614, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224093
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002933 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2416
[LightGBM] [Info] Number of data points in the train set: 18615, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.223451
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
```

```
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002697 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2410
[LightGBM] [Info] Number of data points in the train set: 18615, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.223494
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.005043 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2421
[LightGBM] [Info] Number of data points in the train set: 23268, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224190
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
RMSLE score of stacked models on full data: 0.05974380733376842
```

In [39]:

```python
stacked_model3 = stackedv3.fit(np.array(X_train), np.array(y_train))

#RMSLE score of the stacked model on full train data
stacked_score = rmsle(y_train, stacked_model3.predict(X_train))
print("RMSLE score of stacked models on full data:", stacked_score)
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002355 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2410
[LightGBM] [Info] Number of data points in the train set: 18614, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.222688
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002771 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2417
[LightGBM] [Info] Number of data points in the train set: 18614, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224465
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.004222 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2407
[LightGBM] [Info] Number of data points in the train set: 18614, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.223084
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.002288 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2412
[LightGBM] [Info] Number of data points in the train set: 18615, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.225249
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
```

```
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.005291 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2415
[LightGBM] [Info] Number of data points in the train set: 18615, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.225463
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.006023 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 4221
[LightGBM] [Info] Number of data points in the train set: 23268, number of used feat
ures: 34
[LightGBM] [Info] Start training from score 0.224190
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing wa
s 0.003799 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2421
[LightGBM] [Info] Number of data points in the train set: 23268, number of used feat
ures: 25
[LightGBM] [Info] Start training from score 0.224190
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
```

```
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
RMSLE score of stacked models on full data: 0.053457765494545825
```

In [40]:
```
y_pred = stacked_model.predict(X_test)
y_pred2 = stacked_model2.predict(X_test)
y_pred3 = stacked_model3.predict(X_test)
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
```

In [59]:
```
df_last = pd.read_csv('analysis.csv')
df_last = df_last.tail(2208)
df_last["Pred"] = y_pred
df_test = pd.read_csv("dataset/sample_submission.csv")
a = df_test.copy()
# Ensure both Timestamp columns are in datetime format
df_last['Timestamp'] = pd.to_datetime(df_last['Timestamp'])
df_test['Timestamp'] = pd.to_datetime(df_test['Timestamp'], format='%b %d, %Y %I%p'

# Filter df_last based on the Timestamps in df_test
filtered_df_last = df_last[df_last['Timestamp'].isin(df_test['Timestamp'])]

# Display the filtered dataframe
pred = filtered_df_last['Pred']
a['% Baseline'] = list(pred)
a
```

Out[59]:

	Timestamp	% Baseline
0	Oct 1, 2017 6am	0.043710
1	Oct 1, 2017 7am	0.067176
2	Oct 1, 2017 8am	0.126536
3	Oct 1, 2017 9am	0.317634
4	Oct 1, 2017 10am	0.543645
...	...	...
1072	Dec 31, 2017 12pm	0.289104
1073	Dec 31, 2017 1pm	0.257863
1074	Dec 31, 2017 2pm	0.198282
1075	Dec 31, 2017 3pm	0.216348
1076	Dec 31, 2017 4pm	0.104543

1077 rows × 2 columns

In [60]:
```
a.to_csv('submit/stack_newv1.csv',index=False)
```

In [61]:
```
df_last = pd.read_csv('analysis.csv')
df_last = df_last.tail(2208)
df_last["Pred"] = y_pred2
df_test = pd.read_csv("dataset/sample_submission.csv")
a = df_test.copy()
# Ensure both Timestamp columns are in datetime format
df_last['Timestamp'] = pd.to_datetime(df_last['Timestamp'])
df_test['Timestamp'] = pd.to_datetime(df_test['Timestamp'], format='%b %d, %Y %I%p'

# Filter df_last based on the Timestamps in df_test
filtered_df_last = df_last[df_last['Timestamp'].isin(df_test['Timestamp'])]

# Display the filtered dataframe
pred = filtered_df_last['Pred']
a['% Baseline'] = list(pred)
a
```

Out[61]:

	Timestamp	% Baseline
0	Oct 1, 2017 6am	0.050287
1	Oct 1, 2017 7am	0.057418
2	Oct 1, 2017 8am	0.105584
3	Oct 1, 2017 9am	0.304366
4	Oct 1, 2017 10am	0.507417
...	...	...
1072	Dec 31, 2017 12pm	0.307661
1073	Dec 31, 2017 1pm	0.278214
1074	Dec 31, 2017 2pm	0.122669
1075	Dec 31, 2017 3pm	0.131281
1076	Dec 31, 2017 4pm	0.061643

1077 rows × 2 columns

In [62]:
```
a.to_csv('submit/stack_newv2.csv',index=False)
```

In [63]:
```
df_last = pd.read_csv('analysis.csv')
df_last = df_last.tail(2208)
df_last["Pred"] = y_pred3
df_test = pd.read_csv("dataset/sample_submission.csv")
a = df_test.copy()
# Ensure both Timestamp columns are in datetime format
df_last['Timestamp'] = pd.to_datetime(df_last['Timestamp'])
df_test['Timestamp'] = pd.to_datetime(df_test['Timestamp'], format='%b %d, %Y %I%p'

# Filter df_last based on the Timestamps in df_test
filtered_df_last = df_last[df_last['Timestamp'].isin(df_test['Timestamp'])]

# Display the filtered dataframe
pred = filtered_df_last['Pred']
a['% Baseline'] = list(pred)
a
```

Out[63]:

	Timestamp	% Baseline
0	Oct 1, 2017 6am	0.035036
1	Oct 1, 2017 7am	0.062244
2	Oct 1, 2017 8am	0.117963
3	Oct 1, 2017 9am	0.319705
4	Oct 1, 2017 10am	0.554066
...	...	...
1072	Dec 31, 2017 12pm	0.256863
1073	Dec 31, 2017 1pm	0.203582
1074	Dec 31, 2017 2pm	0.117094
1075	Dec 31, 2017 3pm	0.062047
1076	Dec 31, 2017 4pm	0.045001

1077 rows × 2 columns

In [64]:
```
a.to_csv('submit/stack_newv3.csv',index=False)
```

In [65]:
```
import joblib

# Save the models to disk
joblib.dump(stacked, 'stacked_metaxgb_model.pkl')
joblib.dump(stackedv2, 'stackedv2_metamlp_model.pkl')
joblib.dump(stackedv3, 'stackedv3_metalgb_model.pkl')
```

Out[65]:
```
['stackedv3_metalgb_model.pkl']
```

In [ ]:
```
def blend_models_predict(X):
    return ((0.1 * gbr_model.predict(X)) + \
            (0.125 * rf_model.predict(X)) + \
            (0.05 * en_model.predict(X)) + \
            (0.05 * svr_model.predict(X)) + \
            (0.1 * xgb_model.predict(X)) + \
            (0.15 * lgbm_model.predict(X)) + \
            (0.05 * ridge_model.predict(X)) + \
            (0.05 * lasso_model.predict(X))+ \
            (0.125 * mlp_model.predict(X))+ \
            (0.2 * stacked_model.predict(X)))
```

In [ ]:
```
blend_score = rmsle(y_train, blend_models_predict(X_train))
print("RMSLE score of stacked models on full data:", stacked_score)
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
RMSLE score of stacked models on full data: 0.05089410100385619
```

In [ ]:
```
y_pred4 = blend_models_predict(X_test)
df_test2 = pd.read_csv("dataset/sample_submission.csv")
df_test2["% Baseline"] = y_pred2
```

```
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
[LightGBM] [Warning] min_sum_hessian_in_leaf is set=11, min_child_weight=0.001 will
be ignored. Current value: min_sum_hessian_in_leaf=11
[LightGBM] [Warning] bagging_fraction is set=0.7, subsample=1.0 will be ignored. Cur
rent value: bagging_fraction=0.7
[LightGBM] [Warning] bagging_freq is set=4, subsample_freq=0 will be ignored. Curren
t value: bagging_freq=4
```

In [ ]:
```
df_test2.to_csv('submit/stack_newv5.csv',index=False)
```

In [ ]:
```
```