Forecasting_NG_Demand_Supply

May 3, 2021

1 Forcasting Natural Gas(NG) Demand/Supply

```
[1]: | pip3 install cloudmesh-common
   Collecting cloudmesh-common
     Downloading https://files.pythonhosted.org/packages/08/b9/60e838cd76b05e
   1991ffed2d1387c461a2fefd1e0aa09b230bff0624ff69/cloudmesh_common-4.3.65-py2.py3
   -none-any.whl (80kB)
        || 81kB 3.5MB/s
   Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-
   packages (from cloudmesh-common) (4.41.1)
   Requirement already satisfied: python-dateutil in /usr/local/lib/python3.7/dist-
   packages (from cloudmesh-common) (2.8.1)
   Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-
   packages (from cloudmesh-common) (2.23.0)
   Requirement already satisfied: psutil in /usr/local/lib/python3.7/dist-packages
   (from cloudmesh-common) (5.4.8)
   Collecting pyfiglet
     Downloading https://files.pythonhosted.org/packages/33/07/fcfdd7a2872f5b
   348953de35acce1544dab0c1e8368dca54279b1cde5c15/pyfiglet-0.8.post1-py2.py3-none-
   any.whl (865kB)
        || 870kB 10.5MB/s
   Requirement already satisfied: pathlib in /usr/local/lib/python3.7/dist-
   packages (from cloudmesh-common) (1.0.1)
   Requirement already satisfied: tabulate in /usr/local/lib/python3.7/dist-
   packages (from cloudmesh-common) (0.8.9)
   Collecting python-hostlist
     Downloading https://files.pythonhosted.org/packages/2b/4f/f31dd4b4bf1a57a5c295
   99e1165d0df70dbdddcfa59a7c1d04ee2ff4ccbd/python-hostlist-1.21.tar.gz
   Collecting simplejson
     Downloading https://files.pythonhosted.org/packages/a8/04/377418ac1e530c
   e2a196b54c6552c018fdf1fe776718053efb1f216bffcd/simplejson-3.17.2-cp37-cp37m-
   manylinux2010_x86_64.whl (128kB)
        || 133kB 34.4MB/s
   Collecting oyaml
     Downloading https://files.pythonhosted.org/packages/37/aa/111610d8bf5b1bb7a295
```

```
a048fc648cec346347a8b0be5881defd2d1b4a52/oyaml-1.0-py2.py3-none-any.whl
Collecting colorama
  Downloading https://files.pythonhosted.org/packages/44/98/5b86278fbbf250d239ae
Oecb724f8572af1c91f4a11edf4d36a206189440/colorama-0.4.4-py2.py3-none-any.whl
Requirement already satisfied: humanize in /usr/local/lib/python3.7/dist-
packages (from cloudmesh-common) (0.5.1)
Requirement already satisfied: pytz in /usr/local/lib/python3.7/dist-packages
(from cloudmesh-common) (2018.9)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-
packages (from python-dateutil->cloudmesh-common) (1.15.0)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7
/dist-packages (from requests->cloudmesh-common) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in
/usr/local/lib/python3.7/dist-packages (from requests->cloudmesh-common)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7
/dist-packages (from requests->cloudmesh-common) (2020.12.5)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-
packages (from requests->cloudmesh-common) (2.10)
Requirement already satisfied: pyyaml in /usr/local/lib/python3.7/dist-packages
(from oyaml->cloudmesh-common) (3.13)
Building wheels for collected packages: python-hostlist
  Building wheel for python-hostlist (setup.py) ... done
  Created wheel for python-hostlist: filename=python_hostlist-1.21-cp37-none-
any.whl size=38931
\verb|sha| 256 = \verb|fffc| 1678b9f545997dc94696eb6bff8f68bb638fffe04d5b79bbf06da0be7038| \\
  Stored in directory: /root/.cache/pip/wheels/0b/5b/5b/ddcf52288f0b10f4564ca1b2
531594ff7ccc65f487ba8dc437
Successfully built python-hostlist
Installing collected packages: pyfiglet, python-hostlist, simplejson, oyaml,
colorama, cloudmesh-common
Successfully installed cloudmesh-common-4.3.65 colorama-0.4.4 oyaml-1.0
pyfiglet-0.8.post1 python-hostlist-1.21 simplejson-3.17.2
```

```
[2]: from cloudmesh.common.StopWatch import StopWatch from cloudmesh.common.Shell import Shell

import numpy as np import pandas as pd import matplotlib.pyplot as plt import tensorflow as tf

from tensorflow.keras.models import Sequential from tensorflow.keras.layers import Dense, Activation, SimpleRNN, InputLayer, LSTM, Dropout from tensorflow.keras.utils import to_categorical, plot_model from tensorflow.keras.datasets import mnist
```

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn import metrics
```

1.1 Data Download

Reference: Using Shell.download - cloudmesh

```
[3]: file_url_1 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
     →project/datasets/NaturalGas_Supply_per_Region.csv?raw=true'
   file_url_2 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
    →project/datasets/Tem_korea.csv?raw=true'
   file url 3 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
     →project/datasets/Tem_seoul.csv?raw=true'
   file url 4 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
     →project/datasets/Tem_daegu.csv?raw=true'
   file_url_5 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
    →project/datasets/Tem_busan.csv?raw=true'
   file url 6 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
    →project/datasets/precipitation.csv?raw=true'
   file_url_7 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
     →project/datasets/Pre_seoul.csv?raw=true'
   file_url_8 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
    →project/datasets/Pre_daegu.csv?raw=true'
   file_url_9 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
     →project/datasets/Pre_busan.csv?raw=true'
   file_url_10 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
     →project/datasets/Crude Oil Price.csv?raw=true'
   file_url_11 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
    →project/datasets/Coal_CIF_ARA.xls?raw=true'
   file_url_12 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
     →project/datasets/Coal_Kalimantan.xls?raw=true'
   file url 13 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
     →project/datasets/Coal_Richards_Bay.xls?raw=true'
   file_url_14 = 'https://github.com/cybertraining-dsc/sp21-599-356/blob/main/
    →project/datasets/exchangerate.csv?raw=true'
   destination_1 = '/content/sample_data/NaturalGas_Supply_per_Region.csv'
   destination 2 = '/content/sample data/Tem korea.csv'
   destination_3 = '/content/sample_data/Tem_seoul.csv'
   destination_4 = '/content/sample_data/Tem_daegu.csv'
   destination_5 = '/content/sample_data/Tem_busan.csv'
   destination_6 = '/content/sample_data/precipitation.csv'
   destination_7 = '/content/sample_data/Pre_seoul.csv'
   destination_8 = '/content/sample_data/Pre_daegu.csv'
```

```
destination_9 = '/content/sample_data/Pre_busan.csv'
destination_10 = '/content/sample_data/Crude_Oil_Price.csv'
destination_11 = '/content/sample_data/Coal_CIF_ARA.xls'
destination_12 = '/content/sample_data/Coal_Kalimantan.xls'
destination_13 = '/content/sample_data/Coal_Richards_Bay.xls'
destination_14 = '/content/sample_data/exchangerate.csv'
Shell.download(file_url_1, destination_1, provider='system')
Shell.download(file_url_2, destination_2, provider='system')
Shell.download(file_url_3, destination_3, provider='system')
Shell.download(file_url_4, destination_4, provider='system')
Shell.download(file_url_5, destination_5, provider='system')
Shell.download(file_url_6, destination_6, provider='system')
Shell.download(file_url_7, destination_7, provider='system')
Shell.download(file_url_8, destination_8, provider='system')
Shell.download(file_url_9, destination_9, provider='system')
Shell.download(file_url_10, destination_10, provider='system')
Shell.download(file_url_11, destination_11, provider='system')
Shell.download(file_url_12, destination_12, provider='system')
Shell.download(file_url_13, destination_13, provider='system')
Shell.download(file_url_14, destination_14, provider='system')
```

```
INFO: Used wget
```

[3]: '/content/sample_data/exchangerate.csv'

1.2 Data Pre-Process

1.2.1 Dataset Load

Load the dataset. Each dataset includes time-based monthly data. 1. The amount of natural gas supply (nine cities seperately) 2. The temperature (one national and three regional cities) 3. The precipitation (one national and three regional cities) 4. The price of crude oil (4-types) 5. The price of coal (3-types) 6. The exchange rate between US Dollars(USD) and South Korea Won(KRW)

```
[4]: StopWatch.start("data-load")
   ## Natural gas dataset
   ng sup df = pd.read csv('sample data/NaturalGas Supply per Region.csv')
   #ng pro df = pd.read csv('sample data/NaturalGas Production.csv',,,
    \rightarrow encoding='CP949')
   ## Monthly average temperature dataset
   tem_total = pd.read csv('sample_data/Tem_korea.csv', encoding='CP949')
   tem_df1 = pd.read_csv('sample_data/Tem_seoul.csv', encoding='CP949')
   #tem_df2 = pd.read_csv('sample_data/Tem_Incheon.csv', encoding='CP949')
   #tem_df3 = pd.read csv('sample_data/Tem_suwon_kyunqqi.csv', encodinq='CP949')
   #tem_df4 = pd.read_csv('sample_data/Tem_wonju_gangwon.csv', encoding='CP949')
   #tem df5 = pd.read csv('sample data/Tem daejeon.csv', encoding='CP949')
   #tem_df6 = pd.read_csv('sample_data/Tem_qunsan_jeonbuk.csv', encoding='CP949')
   #tem df7 = pd.read csv('sample data/Tem qwanqju.csv', encoding='CP949')
   tem df8 = pd.read csv('sample data/Tem daegu.csv', encoding='CP949')
   tem df9 = pd.read csv('sample data/Tem busan.csv', encoding='CP949')
   ## Monthly avarage precipation dataset
   pre_total = pd.read_csv('sample_data/precipitation.csv', encoding='CP949')
   pre_df1 = pd.read_csv('sample_data/Pre_seoul.csv', encoding='CP949')
   #pre_df2 = pd.read_csv('sample_data/Pre_incheon.csv', encoding='CP949')
   #pre_df3 = pd.read_csv('sample_data/Pre_suwon_kyungqi.csv', encoding='CP949')
   #pre_df4 = pd.read csv('sample_data/Pre_wonju_qanqwon.csv', encoding='CP949')
   #pre_df5 = pd.read_csv('sample_data/Pre_daejeon.csv', encoding='CP949')
   #pre_df6 = pd.read_csv('sample_data/Pre_qunsan_jeonbuk.csv', encoding='CP949')
   #pre_df7 = pd.read_csv('sample_data/Pre_gwangju.csv', encoding='CP949')
   pre df8 = pd.read csv('sample data/Pre daegu.csv', encoding='CP949')
   pre_df9 = pd.read_csv('sample_data/Pre_busan.csv', encoding='CP949')
   ## Crude oil price dataset
   oil df = pd.read csv('sample data/Crude Oil Price.csv')
   ## Coal price dataset
   coal_df1 = pd.read_excel('sample_data/Coal_CIF_ARA.xls')
   coal_df2 = pd.read_excel('sample_data/Coal_Kalimantan.xls')
   coal_df3 = pd.read_excel('sample_data/Coal_Richards_Bay.xls')
   ## Exchange rate dataset
   ex_df1 = pd.read_csv('sample_data/exchangerate.csv', encoding='CP949')
   StopWatch.stop("data-load")
```

1.2.2 Data preprocess_1

Preprocess data using pandas and create objective dataset.

```
[5]: StopWatch.start("data-preprocess_1")
    ## Change column names to lowercase
    df_names = [ng_sup_df, oil_df, coal_df1, coal_df2, coal_df3]
    for name in df names:
      name.columns = name.columns.str.lower()
    ## Split natural gas dataset
    total_ng_df = ng_sup_df.iloc[:,2:]
    ## Consolidate temperature dataset
    tem_df = pd.concat([tem_df1['avg_tem'],
                        tem_df8['avg_tem'],
                        tem_df9['avg_tem']],axis=1)
    tem_df.columns = ['avg_tem_seoul',
                      'avg_tem_daegu',
                      'avg_tem_busan']
    ## Consolidate precipitation dataset
    pre_df = pd.concat([pre_df1['avg_precipitation'],
                        pre_df8['avg_precipitation'],
                        pre_df9['avg_precipitation']],axis=1)
    pre_df.columns = ['avg_pre_seoul',
                      'avg_pre_daegu',
                      'avg_pre_busan']
    ## Consolidate coal price dataset
    coal_df = pd.
    -concat([coal_df1['price'],coal_df2['price'],coal_df3['price']],axis=1)
    coal_df.columns = ['coal_price_ca', 'coal_price_ka', 'coal_price_rb']
    ## Change exchange rate dataset shape
    ex_df = ex_df1.transpose()
    ex_df.columns = ['rate']
    ex_df.index = list(range(48))
    ## Build objective dataset
    df_total = pd.concat([ng_sup_df['seoul'], tem_df['avg_tem_seoul'],
                          pre_df['avg_pre_seoul'],
                          oil_df[['dubai','brent','wti','oman']],
                          coal_df, ex_df['rate']], axis=1)
    df_total = df_total.iloc[:-1,:]
    StopWatch.stop("data-preprocess_1")
[7]: df_total.head()
[7]:
           seoul avg_tem_seoul ... coal_price_rb
                                                        rate
    0 1110948.0
                           -3.2 ...
                                              40.02 1201.67
       911323.0
                           0.2 ...
                                              42.06 1217.35
```

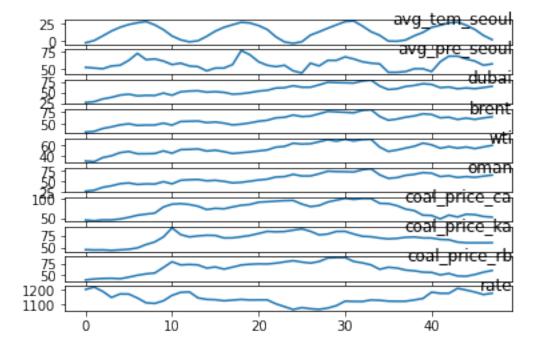
```
      2
      718859.0
      7.0
      ...
      43.16
      1188.21

      3
      417299.0
      14.1
      ...
      43.55
      1147.51

      4
      354428.0
      19.6
      ...
      42.58
      1171.51
```

[5 rows x 11 columns]

```
[8]: ## Plot each column data
fig = plt.figure()
for i in range(1,df_total.shape[-1]):
    ax = fig.add_subplot(df_total.shape[-1]-1,1,i)
    ax.plot(df_total.iloc[:,i])
    ax.set_title(df_total.columns[i], y=0.5, loc='right')
plt.show()
```



```
[9]: ## Make function which can convert normalized dataset to supervised dataset
def convert_dataset(dataset, num_i, num_o, dropnan=True):
    columns = []
    col_names = []
    conv_df = pd.DataFrame(dataset)
    for i in range(num_i, 0, -1):
        columns.append(conv_df.shift(periods=i))
        for j in range(dataset.shape[-1]):
        col_names.append('column{}(t-{})'.format(j+1,i))
        for i in range(num_o):
        columns.append(conv_df.shift(periods=-i))
        if i == 0:
```

```
for j in range(dataset.shape[-1]):
             col_names.append('column{}(t)'.format(j+1))
         else:
           for j in range(dataset.shape[-1]):
             col_names.append('column{}(t+{})'.format(j+1, i))
      new df = pd.concat(columns, axis=1)
       new_df.columns = col_names
       if dropnan:
         new_df.dropna(inplace=True)
       return new df
[10]: ## Make function which can make train and test dataset
     def processed_dataset(dataset_norm, time_interval, boundary):
      new_df = convert_dataset(dataset_norm, time_interval, 1)
      new_df.drop(new_df.columns[list(range(dataset_norm.shape[-1]+1,__
      →2*dataset_norm.shape[-1]))], axis=1, inplace=True)
      train = new_df.values[:boundary, :]
       test = new_df.values[boundary:,:]
       x_train, y_train = train[:, :-1], train[:, -1]
       x_{test}, y_{test} = test[:, :-1], test[:,-1]
       x_train = x_train.reshape(x_train.shape[0], 1, x_train.shape[-1])
       x_test = x_test.reshape(x_test.shape[0], 1, x_test.shape[-1])
      return x_train, y_train, x_test, y_test
[11]: ## Make function which can build the network model
     def define_model(x_train, dropout, learning_rate):
      model = Sequential()
       ## LSTM Layers
      model.add(LSTM(100, input_shape = (x_train.shape[1], x_train.shape[2]),
                      return sequences=True))
       model.add(LSTM(100, dropout=dropout,
                      return_sequences=False))
       ## MLP Layers
       model.add(Dense(100))
      model.add(Activation('relu'))
       model.add(Dropout(dropout))
      model.add(Dense(100))
       model.add(Activation('relu'))
       model.add(Dropout(dropout))
      model.add(Dense(1))
       model.add(Activation('relu'))
```

```
opt = tf.keras.optimizers.Adam(learning_rate=learning_rate)
       model.compile(loss='mae', optimizer=opt)
       model.summary()
       return model
[12]: ## Make function which can train model
     def train_model(model, x_train, y_train, x_test, y_test, epoch):
       history = model.fit(x_train, y_train,
                         epochs=epoch,
                         validation_data=(x_test, y_test))
       return history
[13]: ## Make function which can plot loss
     def loss_plot(history):
      k = list(range(1, len(history.history['loss'])+1))
       plt.plot(k, history.history['loss'], label='Train Loss')
      plt.plot(k, history.history['val_loss'], label='Validation Loss')
      plt.ylabel('Mean Absolute Error')
      plt.xlabel('Epoch')
      plt.legend()
      return plt.show()
[14]: | ## Make function which can convert dataset to original shape
     def predicted_model(model, x_test, y_test, scaler):
      y_predicted = model.predict(x_test)
       x_test = x_test.reshape(x_test.shape[0], x_test.shape[2])
      x_te_re = x_test[:,1:]
       y_test = y_test.reshape(len(y_test), 1)
       inv_y_predicted = np.concatenate((y_predicted, x_te_re), axis=1)
       inv_y_predicted = scaler.inverse_transform(inv_y_predicted)[:,0]
       inv_y = np.concatenate((y_test, x_te_re), axis=1)
       inv_y = scaler.inverse_transform(inv_y)[:,0]
       rmse = np.sqrt(metrics.mean_squared_error(inv_y, inv_y_predicted))
       print('The RMSE is: %.4f' % rmse)
      plt.plot(inv_y, label='Real')
      plt.plot(inv_y_predicted, label='Prediction')
      plt.legend()
       return plt.show()
[15]: | ## Make function which can convert dataset to original shape(with timesteps)
     def predicted_time_model(model, x_test, y_test, scaler, time, feature):
       y_predicted = model.predict(x_test)
```

```
x_test = x_test.reshape(x_test.shape[0], (time*feature))
x_te_re = x_test[:,-(feature-1):]
y_test = y_test.reshape(len(y_test), 1)

inv_y_predicted = np.concatenate((y_predicted, x_te_re), axis=1)
inv_y_predicted = scaler.inverse_transform(inv_y_predicted)[:,0]

inv_y = np.concatenate((y_test, x_te_re), axis=1)
inv_y = scaler.inverse_transform(inv_y)[:,0]

rmse = np.sqrt(metrics.mean_squared_error(inv_y, inv_y_predicted))
print('The RMSE is : %.4f' % rmse)

plt.plot(inv_y, label='Real')
plt.plot(inv_y_predicted, label='Prediction')
plt.legend()

return plt.show()
```

1.2.3 Data preprocess_2

Preprocess data using various function. Data can be normalized and divided into train and test set

Reference: How to use MinMaxScaler

Reference: How to process time series dataset

Reference: How to process multivariate time series dataset

```
[16]: StopWatch.start("data-preprocess_2")
    scaler = MinMaxScaler()
    dataset_norm = scaler.fit_transform(df_total)
    x_train, y_train, x_test, y_test = processed_dataset(dataset_norm, 1, 12)
    StopWatch.stop("data-preprocess_2")
```

1.3 Define Model

Build the network model.

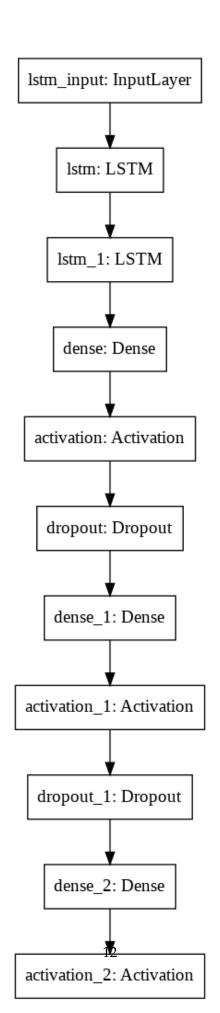
Reference: Introduction to MAE and RMSE Reference: MLP + LSTM with Tensorflow

```
[17]: StopWatch.start("compile")
  model = define_model(x_train, 0.1, 0.0005)
  StopWatch.stop("compile")
  tf.keras.utils.plot_model(model)
```

lstm (LSTM)	(None, 1, 100)	44800
lstm_1 (LSTM)	(None, 100)	80400
dense (Dense)	(None, 100)	10100
activation (Activation)	(None, 100)	0
dropout (Dropout)	(None, 100)	0
dense_1 (Dense)	(None, 100)	10100
activation_1 (Activation)	(None, 100)	0
dropout_1 (Dropout)	(None, 100)	0
dense_2 (Dense)	(None, 1)	101
activation_2 (Activation)	(None, 1)	0

Total params: 145,501 Trainable params: 145,501 Non-trainable params: 0

[17]:



1.4 Train

Epoch 13/100

```
Train the dataset.
[18]: StopWatch.start("train")
 history = train model(model, x_train, y_train, x_test, y_test, 100)
 StopWatch.stop("train")
 Epoch 1/100
 0.3747
 Epoch 2/100
 0.3664
 Epoch 3/100
 0.3583
 Epoch 4/100
 0.3498
 Epoch 5/100
 0.3409
 Epoch 6/100
 0.3317
 Epoch 7/100
 0.3226
 Epoch 8/100
 0.3132
 Epoch 9/100
 0.3044
 Epoch 10/100
 0.2971
 Epoch 11/100
 0.2909
 Epoch 12/100
 0.2851
```

```
0.2797
Epoch 14/100
0.2747
Epoch 15/100
0.2701
Epoch 16/100
0.2656
Epoch 17/100
0.2613
Epoch 18/100
0.2581
Epoch 19/100
0.2551
Epoch 20/100
0.2521
Epoch 21/100
0.2500
Epoch 22/100
0.2490
Epoch 23/100
0.2487
Epoch 24/100
0.2491
Epoch 25/100
0.2487
Epoch 26/100
0.2476
Epoch 27/100
0.2464
Epoch 28/100
0.2448
Epoch 29/100
```

```
0.2439
Epoch 30/100
0.2427
Epoch 31/100
0.2405
Epoch 32/100
0.2379
Epoch 33/100
0.2348
Epoch 34/100
0.2309
Epoch 35/100
0.2263
Epoch 36/100
0.2211
Epoch 37/100
0.2157
Epoch 38/100
0.2123
Epoch 39/100
0.2092
Epoch 40/100
0.2060
Epoch 41/100
0.2023
Epoch 42/100
0.1972
Epoch 43/100
0.1931
Epoch 44/100
0.1888
Epoch 45/100
```

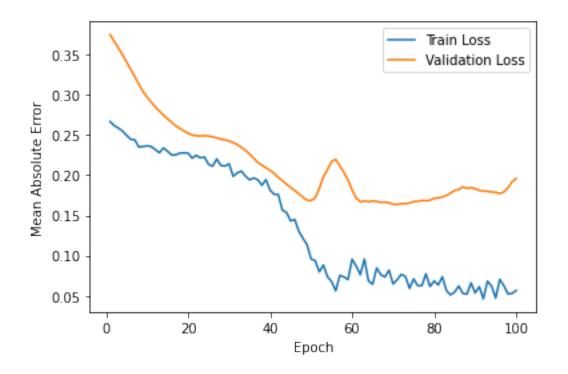
```
0.1848
Epoch 46/100
0.1810
Epoch 47/100
0.1768
Epoch 48/100
0.1727
Epoch 49/100
0.1689
Epoch 50/100
0.1684
Epoch 51/100
0.1718
Epoch 52/100
0.1836
Epoch 53/100
0.1980
Epoch 54/100
0.2068
Epoch 55/100
0.2171
Epoch 56/100
0.2196
Epoch 57/100
0.2117
Epoch 58/100
0.2036
Epoch 59/100
0.1939
Epoch 60/100
0.1816
Epoch 61/100
```

```
0.1717
Epoch 62/100
0.1669
Epoch 63/100
0.1679
Epoch 64/100
0.1671
Epoch 65/100
0.1680
Epoch 66/100
0.1675
Epoch 67/100
0.1661
Epoch 68/100
0.1665
Epoch 69/100
0.1657
Epoch 70/100
0.1637
Epoch 71/100
0.1637
Epoch 72/100
0.1647
Epoch 73/100
0.1645
Epoch 74/100
0.1654
Epoch 75/100
0.1673
Epoch 76/100
0.1674
Epoch 77/100
```

```
0.1686
Epoch 78/100
0.1683
Epoch 79/100
0.1688
Epoch 80/100
0.1714
Epoch 81/100
0.1718
Epoch 82/100
0.1729
Epoch 83/100
0.1749
Epoch 84/100
0.1777
Epoch 85/100
0.1811
Epoch 86/100
0.1822
Epoch 87/100
0.1855
Epoch 88/100
0.1837
Epoch 89/100
0.1847
Epoch 90/100
0.1830
Epoch 91/100
0.1811
Epoch 92/100
0.1802
Epoch 93/100
```

```
0.1801
Epoch 94/100
0.1790
Epoch 95/100
1/1 [=====
              ======] - Os 101ms/step - loss: 0.0474 - val_loss:
0.1788
Epoch 96/100
              ======] - Os 101ms/step - loss: 0.0704 - val_loss:
1/1 [=====
0.1771
Epoch 97/100
0.1792
Epoch 98/100
0.1844
Epoch 99/100
0.1917
Epoch 100/100
1/1 [======
            =======] - Os 92ms/step - loss: 0.0566 - val_loss:
0.1960
```

[19]: loss_plot(history)

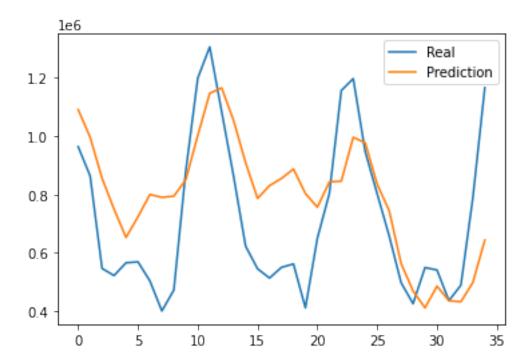


1.5 Predict

Predict from training and compare with original dataset.

```
[20]: StopWatch.start("predict")
    predicted_model(model, x_test, y_test, scaler)
    StopWatch.stop("predict")
```

The RMSE is: 227017.8431



Reference: StopWatch and Benchmark

```
[21]: StopWatch.benchmark()
```

```
| DISTRIB_ID
                 | Ubuntu
                 | 18.04
| DISTRIB_RELEASE
| HOME URL
                 | "https://www.ubuntu.com/"
                  | ubuntu
| ID
| ID_LIKE
                 debian
                 | "Ubuntu"
| NAME
                 | "Ubuntu 18.04.5 LTS"
| PRETTY_NAME
| PRIVACY_POLICY_URL | "https://www.ubuntu.com/legal/terms-and-policies
/privacy-policy" |
| SUPPORT_URL
                  | "https://help.ubuntu.com/"
| UBUNTU_CODENAME
                  bionic
                  | "18.04.5 LTS (Bionic Beaver)"
| VERSION
| VERSION_CODENAME
                 bionic
| VERSION_ID
                 | "18.04"
                 1 2
| cpu_count
| mem.active
                 | 1.1 GiB
| mem.available
                 | 11.7 GiB
| mem.free
                 | 9.9 GiB
| mem.inactive
                 | 1.4 GiB
| mem.percent
                 8.0 %
| mem.total
                 | 12.7 GiB
                 | 792.3 MiB
mem.used
| python
                 | 3.7.10 (default, Feb 20 2021, 21:17:23)
                  | [GCC 7.5.0]
```

```
| python.pip
               | 19.3.1
               | 3.7.10
| python.version
| sys.platform
               linux
| uname.machine
               | x86 64
               | 554b9d81d372
| uname.node
uname.processor
               | x86_64
               | 4.19.112+
| uname.release
| uname.system
               | Linux
 uname.version
               | #1 SMP Thu Jul 23 08:00:38 PDT 2020
user
               | collab
+----+
______
              | Status
                      1
                        Time |
                                                   | tag
| Name
                                Sum | Start
               I OS
Node
         | User
                     | Version
|-----
------
| data-load
              l ok
                     | 0.093 | 0.093 | 2021-05-02 23:09:53 |
554b9d81d372 | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| data-preprocess_1 | ok
                     | 0.022 | 0.022 | 2021-05-02 23:09:55 |
554b9d81d372 | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| data-preprocess_2 | ok
                   | 0.037 | 0.037 | 2021-05-02 23:10:12 |
554b9d81d372 | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
              ok
| compile
                     0.984 | 0.984 | 2021-05-02 23:10:14 |
554b9d81d372 | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
              lok
                     | 16.181 | 16.181 | 2021-05-02 23:11:05 |
554b9d81d372 | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
                     | 1.063 | 1.063 | 2021-05-02 23:11:22 |
              ok
predict
554b9d81d372 | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
+----+
______
```

[#] csv,timer,status,time,sum,start,tag,uname.node,user,uname.system,platform.vers
ion

[#] csv,data-load,ok,0.093,0.093,2021-05-02 23:09:53,,554b9d81d372,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020

```
# csv,data-preprocess_1,ok,0.022,0.022,2021-05-02
23:09:55,,554b9d81d372,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,data-preprocess_2,ok,0.037,0.037,2021-05-02
23:10:12,,554b9d81d372,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,compile,ok,0.984,0.984,2021-05-02 23:10:14,,554b9d81d372,collab,Linux,#1
SMP Thu Jul 23 08:00:38 PDT 2020
# csv,train,ok,16.181,16.181,2021-05-02 23:11:05,,554b9d81d372,collab,Linux,#1
SMP Thu Jul 23 08:00:38 PDT 2020
# csv,predict,ok,1.063,1.063,2021-05-02 23:11:22,,554b9d81d372,collab,Linux,#1
SMP Thu Jul 23 08:00:38 PDT 2020
```

1.6 Test1 - Climate dataset

In this test, the climate dataset is used to build a model. The dataset includes temperature data and precipitation data.

```
[32]:
          seoul avg_tem_seoul avg_pre_seoul
     0 1110948
                           -3.2
                                             53
                                             52
     1
         911323
                            0.2
     2
         718859
                            7.0
                                             51
     3
         417299
                           14.1
                                             55
         354428
                           19.6
```

```
[33]: StopWatch.start("test1-data-preprocess_2")
scaler2 = MinMaxScaler()
dataset_norm2 = scaler2.fit_transform(df_total_ver_2)
x_train2, y_train2, x_test2, y_test2 = processed_dataset(dataset_norm2, 1, 12)
x_train2.shape, y_train2.shape, x_test2.shape, y_test2.shape
StopWatch.stop("test1-data-preprocess_2")
```

```
[34]: StopWatch.start("test2-compile")
model2 = define_model(x_train2, 0.1, 0.0005)
StopWatch.stop("test2-compile")
tf.keras.utils.plot_model(model2)
```

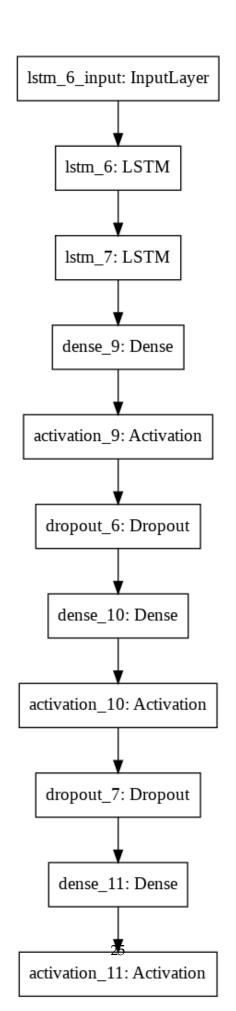
Model: "sequential_3"

Layer (type)	Output Shape	Param #
lstm_6 (LSTM)	(None, 1, 100)	41600
lstm_7 (LSTM)	(None, 100)	80400

dense_9 (Dense)	(None, 100)	10100
activation_9 (Activation)	(None, 100)	0
dropout_6 (Dropout)	(None, 100)	0
dense_10 (Dense)	(None, 100)	10100
activation_10 (Activation)	(None, 100)	0
dropout_7 (Dropout)	(None, 100)	0
dense_11 (Dense)	(None, 1)	101
activation_11 (Activation)	(None, 1)	0

Total params: 142,301 Trainable params: 142,301 Non-trainable params: 0

[34]:



```
[35]: StopWatch.start("test1-train")
 history2 = train_model(model2, x_train2, y_train2, x_test2, y_test2, 100)
 StopWatch.stop("test1-train")
 Epoch 1/100
 0.3820
 Epoch 2/100
 0.3820
 Epoch 3/100
 0.3801
 Epoch 4/100
 0.3760
 Epoch 5/100
 0.3715
 Epoch 6/100
 0.3671
 Epoch 7/100
 0.3627
 Epoch 8/100
 0.3582
 Epoch 9/100
 0.3533
 Epoch 10/100
 0.3482
 Epoch 11/100
 0.3428
 Epoch 12/100
 0.3371
 Epoch 13/100
 0.3312
 Epoch 14/100
```

0.3252

```
Epoch 15/100
0.3192
Epoch 16/100
0.3131
Epoch 17/100
0.3070
Epoch 18/100
0.3012
Epoch 19/100
0.2966
Epoch 20/100
0.2923
Epoch 21/100
0.2884
Epoch 22/100
Epoch 23/100
0.2814
Epoch 24/100
0.2782
Epoch 25/100
0.2752
Epoch 26/100
0.2723
Epoch 27/100
0.2695
Epoch 28/100
0.2669
Epoch 29/100
0.2644
Epoch 30/100
0.2620
```

```
Epoch 31/100
0.2606
Epoch 32/100
0.2590
Epoch 33/100
0.2574
Epoch 34/100
0.2559
Epoch 35/100
0.2542
Epoch 36/100
0.2525
Epoch 37/100
0.2511
Epoch 38/100
0.2496
Epoch 39/100
1/1 [============= ] - Os 91ms/step - loss: 0.2174 - val_loss:
0.2492
Epoch 40/100
0.2485
Epoch 41/100
0.2476
Epoch 42/100
0.2467
Epoch 43/100
0.2457
Epoch 44/100
0.2446
Epoch 45/100
0.2433
Epoch 46/100
0.2411
```

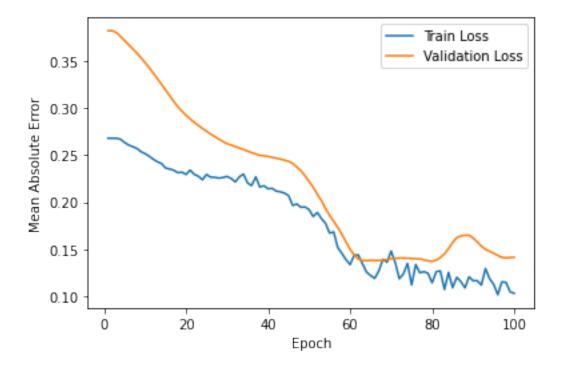
```
Epoch 47/100
0.2373
Epoch 48/100
0.2333
Epoch 49/100
0.2279
Epoch 50/100
0.2218
Epoch 51/100
0.2158
Epoch 52/100
0.2084
Epoch 53/100
0.2016
Epoch 54/100
0.1933
Epoch 55/100
0.1860
Epoch 56/100
0.1796
Epoch 57/100
0.1731
Epoch 58/100
0.1658
Epoch 59/100
0.1576
Epoch 60/100
0.1499
Epoch 61/100
0.1440
Epoch 62/100
0.1395
```

```
Epoch 63/100
0.1380
Epoch 64/100
0.1379
Epoch 65/100
0.1382
Epoch 66/100
0.1381
Epoch 67/100
0.1380
Epoch 68/100
0.1384
Epoch 69/100
0.1391
Epoch 70/100
0.1399
Epoch 71/100
1/1 [============ ] - Os 88ms/step - loss: 0.1359 - val_loss:
0.1404
Epoch 72/100
0.1406
Epoch 73/100
0.1406
Epoch 74/100
0.1405
Epoch 75/100
0.1401
Epoch 76/100
0.1400
Epoch 77/100
0.1396
Epoch 78/100
0.1387
```

```
Epoch 79/100
0.1378
Epoch 80/100
0.1371
Epoch 81/100
0.1385
Epoch 82/100
0.1412
Epoch 83/100
0.1450
Epoch 84/100
0.1508
Epoch 85/100
0.1571
Epoch 86/100
0.1624
Epoch 87/100
0.1641
Epoch 88/100
0.1648
Epoch 89/100
0.1649
Epoch 90/100
0.1619
Epoch 91/100
0.1578
Epoch 92/100
0.1533
Epoch 93/100
0.1498
Epoch 94/100
0.1478
```

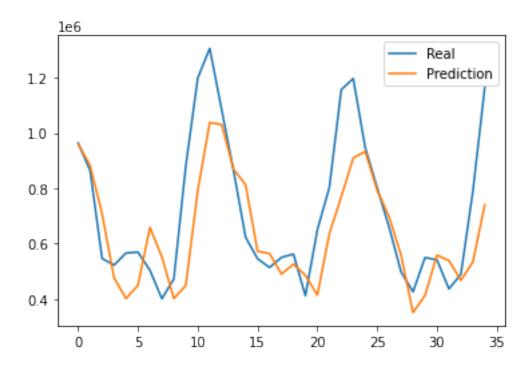
```
Epoch 95/100
0.1457
Epoch 96/100
1/1 [======
       =======] - Os 97ms/step - loss: 0.1017 - val_loss:
0.1433
Epoch 97/100
0.1414
Epoch 98/100
0.1407
Epoch 99/100
0.1411
Epoch 100/100
0.1414
```

[36]: loss_plot(history2)



```
[37]: StopWatch.start("test1-predict")
predicted_model(model2, x_test2, y_test2, scaler2)
StopWatch.stop("test1-predict")
```

The RMSE is: 185204.7026



1.7 Test2 - Temperature dataset

In this test, only a temperature dataset is used to build a model. The values of this dataset show that it is low in winter and high in summer.

```
[41]: StopWatch.start("test2-data-preprocess_1")

df_total_ver_3 = pd.concat([ng_sup_df['seoul'], tem_df['avg_tem_seoul']],

→axis=1)

StopWatch.stop("test2-data-preprocess_1")

df_total_ver_3.head()
```

```
[41]:
           seoul
                  avg_tem_seoul
        1110948
                             -3.2
     1
          911323
                              0.2
     2
          718859
                              7.0
                             14.1
     3
          417299
          354428
                             19.6
```

```
[42]: StopWatch.start("test2-data-preprocess_2")
scaler3 = MinMaxScaler()
dataset_norm3 = scaler3.fit_transform(df_total_ver_3)
x_train3, y_train3, x_test3, y_test3 = processed_dataset(dataset_norm3, 1, 12)
StopWatch.stop("test2-data-preprocess_2")
x_train3.shape, y_train3.shape, x_test3.shape, y_test3.shape
```

```
[42]: ((12, 1, 2), (12,), (35, 1, 2), (35,))
```

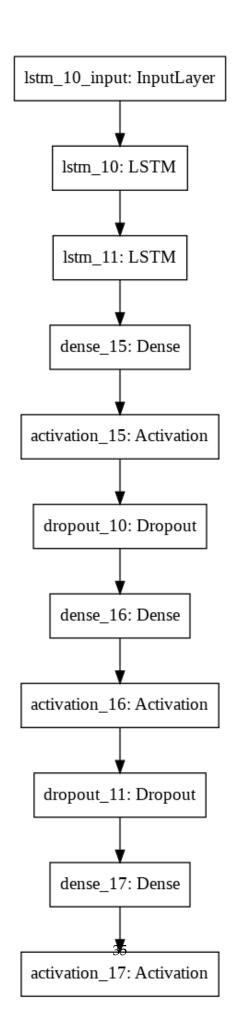
```
[43]: StopWatch.start("test2-compile")
model3 = define_model(x_train3, 0.1, 0.0005)
StopWatch.stop("test2-compile")
tf.keras.utils.plot_model(model3)
```

Model: "sequential_5"

Layer (type)	Output Shape	Param #
lstm_10 (LSTM)	(None, 1, 100)	41200
lstm_11 (LSTM)	(None, 100)	80400
dense_15 (Dense)	(None, 100)	10100
activation_15 (Activation)	(None, 100)	0
dropout_10 (Dropout)	(None, 100)	0
dense_16 (Dense)	(None, 100)	10100
activation_16 (Activation)	(None, 100)	0
dropout_11 (Dropout)	(None, 100)	0
dense_17 (Dense)	(None, 1)	101
activation_17 (Activation)	(None, 1)	0
m . 7		

Total params: 141,901 Trainable params: 141,901 Non-trainable params: 0

[43]:



```
[44]: StopWatch.start("test2-train")
 history3 = train_model(model3, x_train3, y_train3, x_test3, y_test3, 100)
 StopWatch.stop("test2-train")
 Epoch 1/100
 0.3768
 Epoch 2/100
 0.3712
 Epoch 3/100
 0.3661
 Epoch 4/100
 0.3608
 Epoch 5/100
 0.3553
 Epoch 6/100
 0.3494
 Epoch 7/100
 0.3432
 Epoch 8/100
 0.3366
 Epoch 9/100
 0.3297
 Epoch 10/100
 0.3228
 Epoch 11/100
 0.3159
 Epoch 12/100
 0.3089
 Epoch 13/100
 0.3023
 Epoch 14/100
```

0.2962

```
Epoch 15/100
0.2908
Epoch 16/100
0.2858
Epoch 17/100
0.2813
Epoch 18/100
0.2770
Epoch 19/100
0.2729
Epoch 20/100
0.2692
Epoch 21/100
0.2655
Epoch 22/100
0.2629
Epoch 23/100
0.2604
Epoch 24/100
0.2580
Epoch 25/100
0.2568
Epoch 26/100
0.2555
Epoch 27/100
0.2550
Epoch 28/100
0.2552
Epoch 29/100
0.2552
Epoch 30/100
0.2551
```

```
Epoch 31/100
0.2549
Epoch 32/100
0.2546
Epoch 33/100
0.2551
Epoch 34/100
0.2554
Epoch 35/100
0.2556
Epoch 36/100
0.2555
Epoch 37/100
0.2547
Epoch 38/100
0.2540
Epoch 39/100
0.2528
Epoch 40/100
0.2515
Epoch 41/100
0.2498
Epoch 42/100
0.2477
Epoch 43/100
0.2452
Epoch 44/100
0.2401
Epoch 45/100
0.2347
Epoch 46/100
0.2289
```

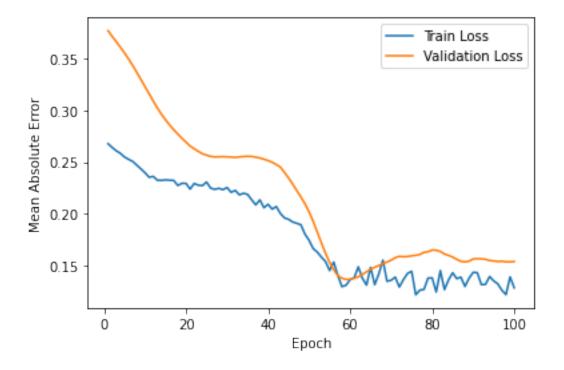
```
Epoch 47/100
0.2225
Epoch 48/100
0.2164
Epoch 49/100
0.2098
Epoch 50/100
0.2020
Epoch 51/100
0.1928
Epoch 52/100
0.1824
Epoch 53/100
0.1722
Epoch 54/100
0.1621
Epoch 55/100
1/1 [============ ] - Os 87ms/step - loss: 0.1454 - val_loss:
0.1534
Epoch 56/100
0.1461
Epoch 57/100
0.1405
Epoch 58/100
0.1381
Epoch 59/100
0.1368
Epoch 60/100
0.1370
Epoch 61/100
0.1380
Epoch 62/100
0.1394
```

```
Epoch 63/100
0.1417
Epoch 64/100
0.1442
Epoch 65/100
0.1465
Epoch 66/100
0.1486
Epoch 67/100
0.1503
Epoch 68/100
0.1520
Epoch 69/100
0.1536
Epoch 70/100
0.1554
Epoch 71/100
1/1 [============ ] - Os 93ms/step - loss: 0.1391 - val_loss:
0.1578
Epoch 72/100
0.1591
Epoch 73/100
0.1588
Epoch 74/100
0.1589
Epoch 75/100
0.1596
Epoch 76/100
0.1601
Epoch 77/100
0.1609
Epoch 78/100
0.1630
```

```
Epoch 79/100
0.1635
Epoch 80/100
0.1652
Epoch 81/100
0.1650
Epoch 82/100
0.1639
Epoch 83/100
0.1614
Epoch 84/100
0.1600
Epoch 85/100
0.1584
Epoch 86/100
0.1562
Epoch 87/100
1/1 [============ ] - Os 92ms/step - loss: 0.1391 - val_loss:
0.1543
Epoch 88/100
0.1538
Epoch 89/100
0.1545
Epoch 90/100
0.1566
Epoch 91/100
0.1568
Epoch 92/100
0.1567
Epoch 93/100
0.1563
Epoch 94/100
0.1551
```

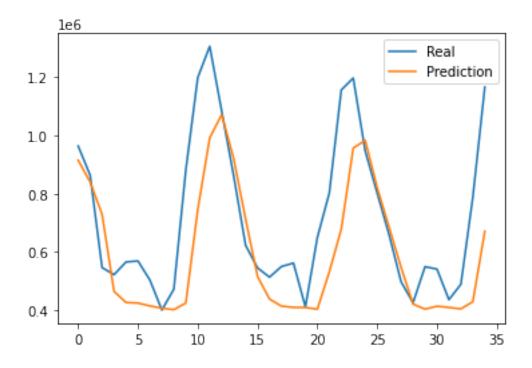
```
Epoch 95/100
0.1547
Epoch 96/100
1/1 [======
       =======] - Os 84ms/step - loss: 0.1326 - val_loss:
0.1541
Epoch 97/100
0.1544
Epoch 98/100
0.1538
Epoch 99/100
0.1538
Epoch 100/100
0.1541
```

[45]: loss_plot(history3)



```
[46]: StopWatch.start("test2-predict")
predicted_model(model3, x_test3, y_test3, scaler3)
StopWatch.stop("test2-predict")
```

The RMSE is: 207585.1521



[120]:

1.8 Test3 - Applying timesteps

In this test, all dataset are used to build a model. The dataset is same to the first implementation, but timesteps are applied.

```
[47]: StopWatch.start("test3-data-preprocess_1")
    df_total_ver_4 = df_total.copy()
    StopWatch.stop("test3-data-preprocess_1")
    df_total_ver_4.head()
```

```
[47]:
            seoul avg_tem_seoul ...
                                       coal_price_rb
                                                          rate
                            -3.2 ...
        1110948.0
                                                40.02 1201.67
         911323.0
                             0.2 ...
                                                42.06
                                                       1217.35
     1
     2
         718859.0
                             7.0 ...
                                                43.16
                                                       1188.21
         417299.0
                            14.1
     3
                                                43.55
                                                       1147.51
                                  . . .
         354428.0
                            19.6
                                                42.58 1171.51
```

[5 rows x 11 columns]

```
[48]: StopWatch.start("test3-data-preprocess_2")
scaler4 = MinMaxScaler()
dataset_norm4 = scaler4.fit_transform(df_total_ver_4)
```

```
months = 2
features = 11
n = months*features

new_df = convert_dataset(dataset_norm4, months, 1)
values = new_df.values
train4 = values[:12, :]
test4 = values[12:,:]

x_train4, y_train4 = train4[:, :n], train4[:, -features]
x_test4, y_test4 = test4[:, :n], test4[:,-features]

x_train4 = x_train4.reshape(x_train4.shape[0], months, features)
x_test4 = x_test4.reshape(x_test4.shape[0], months, features)
StopWatch.stop("test3-data-preprocess_2")
x_train4.shape, y_train4.shape, x_test4.shape, y_test4.shape
```

[48]: ((12, 2, 11), (12,), (34, 2, 11), (34,))

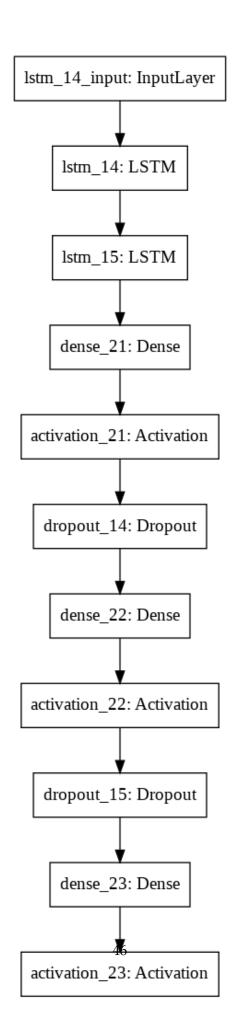
```
[50]: StopWatch.start("test3-compile")
  model4 = define_model(x_train4, 0.1, 0.0005)
  StopWatch.stop("test3-compile")
  tf.keras.utils.plot_model(model4)
```

Model: "sequential_7"

Layer (type)	Output Shape	Param #
lstm_14 (LSTM)	(None, 2, 100)	44800
lstm_15 (LSTM)	(None, 100)	80400
dense_21 (Dense)	(None, 100)	10100
activation_21 (Activation)	(None, 100)	0
dropout_14 (Dropout)	(None, 100)	0
dense_22 (Dense)	(None, 100)	10100
activation_22 (Activation)	(None, 100)	0
dropout_15 (Dropout)	(None, 100)	0
dense_23 (Dense)	(None, 1)	101
activation_23 (Activation)	(None, 1)	0

Total params: 145,501 Trainable params: 145,501 Non-trainable params: 0

[50]:



```
[51]: StopWatch.start("test3-train")
 history4 = train_model(model4, x_train4, y_train4, x_test4, y_test4, 100)
 StopWatch.stop("test3-train")
 Epoch 1/100
 0.3351
 Epoch 2/100
 0.3173
 Epoch 3/100
 0.3009
 Epoch 4/100
 0.2860
 Epoch 5/100
 0.2716
 Epoch 6/100
 0.2590
 Epoch 7/100
 0.2468
 Epoch 8/100
 0.2354
 Epoch 9/100
 0.2270
 Epoch 10/100
 0.2217
 Epoch 11/100
 0.2184
 Epoch 12/100
 0.2160
 Epoch 13/100
 0.2139
 Epoch 14/100
```

0.2119

```
Epoch 15/100
1/1 [============ ] - Os 98ms/step - loss: 0.1998 - val_loss:
0.2099
Epoch 16/100
0.2081
Epoch 17/100
0.2069
Epoch 18/100
0.2054
Epoch 19/100
0.2039
Epoch 20/100
0.2027
Epoch 21/100
0.2014
Epoch 22/100
0.2007
Epoch 23/100
0.2002
Epoch 24/100
0.2009
Epoch 25/100
0.2062
Epoch 26/100
0.2172
Epoch 27/100
0.2350
Epoch 28/100
0.2584
Epoch 29/100
0.2838
Epoch 30/100
0.3184
```

```
Epoch 31/100
0.3451
Epoch 32/100
0.3541
Epoch 33/100
0.3545
Epoch 34/100
0.3485
Epoch 35/100
0.3356
Epoch 36/100
0.3245
Epoch 37/100
0.3193
Epoch 38/100
Epoch 39/100
0.3392
Epoch 40/100
0.3470
Epoch 41/100
0.3659
Epoch 42/100
0.3750
Epoch 43/100
0.3789
Epoch 44/100
0.3676
Epoch 45/100
0.3547
Epoch 46/100
0.3372
```

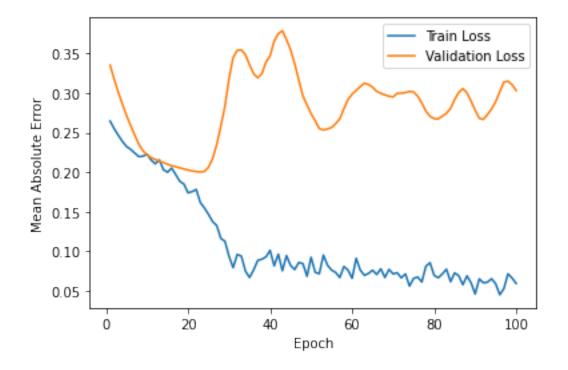
```
Epoch 47/100
0.3164
Epoch 48/100
0.2964
Epoch 49/100
0.2849
Epoch 50/100
0.2738
Epoch 51/100
0.2649
Epoch 52/100
0.2551
Epoch 53/100
0.2533
Epoch 54/100
0.2546
Epoch 55/100
0.2564
Epoch 56/100
0.2614
Epoch 57/100
0.2673
Epoch 58/100
0.2803
Epoch 59/100
0.2919
Epoch 60/100
0.2985
Epoch 61/100
0.3036
Epoch 62/100
0.3082
```

```
Epoch 63/100
1/1 [============ ] - Os 77ms/step - loss: 0.0694 - val_loss:
0.3126
Epoch 64/100
0.3108
Epoch 65/100
0.3074
Epoch 66/100
0.3021
Epoch 67/100
0.2997
Epoch 68/100
0.2976
Epoch 69/100
0.2960
Epoch 70/100
0.2951
Epoch 71/100
0.2997
Epoch 72/100
0.2997
Epoch 73/100
0.3007
Epoch 74/100
0.3021
Epoch 75/100
0.3014
Epoch 76/100
0.2960
Epoch 77/100
0.2874
Epoch 78/100
0.2770
```

```
Epoch 79/100
0.2706
Epoch 80/100
0.2676
Epoch 81/100
0.2674
Epoch 82/100
0.2707
Epoch 83/100
0.2744
Epoch 84/100
0.2806
Epoch 85/100
0.2919
Epoch 86/100
Epoch 87/100
0.3056
Epoch 88/100
0.3003
Epoch 89/100
0.2899
Epoch 90/100
0.2782
Epoch 91/100
0.2680
Epoch 92/100
0.2668
Epoch 93/100
0.2729
Epoch 94/100
0.2799
```

```
Epoch 95/100
0.2892
Epoch 96/100
1/1 [======
          =======] - Os 101ms/step - loss: 0.0448 - val_loss:
0.3016
Epoch 97/100
       ======== ] - Os 103ms/step - loss: 0.0526 - val_loss:
0.3140
Epoch 98/100
0.3148
Epoch 99/100
0.3106
Epoch 100/100
0.3032
```

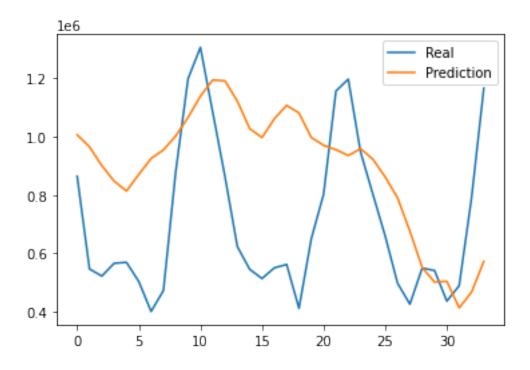
[52]: loss_plot(history4)



[53]: StopWatch.start("test3-predict")
predicted_time_model(model4, x_test4, y_test4, scaler4, months, features)
StopWatch.stop("test3-predict")

WARNING:tensorflow:5 out of the last 9 calls to <function Model.make_predict_function.<locals>.predict_function at 0x7f46a9aae5f0> triggered tf.function retracing. Tracing is expensive and the excessive number of tracings could be due to (1) creating Otf.function repeatedly in a loop, (2) passing tensors with different shapes, (3) passing Python objects instead of tensors. For (1), please define your @tf.function outside of the loop. For (2), @tf.function has experimental relax shapes=True option that relaxes argument shapes that can avoid unnecessary retracing. For (3), please refer to https://www.tensorflow.org/guide/function#controlling_retracing and https://www.tensorflow.org/api_docs/python/tf/function for more details.

The RMSE is : 340842.9452



[155]:

1.9 Test4 - National dataset

```
[54]: StopWatch.start("test4-data-preprocess_1")
     df_total_country = pd.concat([ng_sup_df.sum(axis=1), tem_total['avg_tem'],
                                   pre_total['avg_precipitation'],
                                   oil_df[['dubai','brent','wti','oman']],
                                   coal_df, ex_df['rate']], axis=1)
     df_total_country = df_total_country.iloc[:-1,:]
     df_total_country = df_total_country.rename(columns = {0:'total'})
     StopWatch.stop("test4-data-preprocess 1")
     df_total_country.head()
```

```
[54]:
           total avg_tem avg_precipitation ... coal_price_ka coal_price_rb
    rate
    0 4250103.0
                                                         46.33
                                                                        40.02
                     2.8
                                       26.3 ...
    1201.67
    1 3559794.0
                    8.8
                                       58.2 ...
                                                         45.67
                                                                        42.06
    1217.35
    2 3182961.0
                    15.8
                                      169.0 ...
                                                         45.89
                                                                        43.16
    1188.21
    3 2125051.0
                    21.8
                                      221.2 ...
                                                         44.92
                                                                        43.55
    1147.51
    4 1978337.0
                    26.2
                                                                        42.58
                                      140.0 ...
                                                         45.98
    1171.51
```

[5 rows x 11 columns]

```
[55]: StopWatch.start("test4-data-preprocess_2")
    scaler5 = MinMaxScaler()
    dataset_norm5 = scaler5.fit_transform(df_total_country)
    x_train5, y_train5, x_test5, y_test5 = processed_dataset(dataset_norm5, 1, 12)
    StopWatch.stop("test4-data-preprocess_2")
```

[57]: StopWatch.start("test4-compile")
model5 = define_model(x_train5, 0.1, 0.0005)
StopWatch.stop("test4-compile")
tf.keras.utils.plot_model(model5)

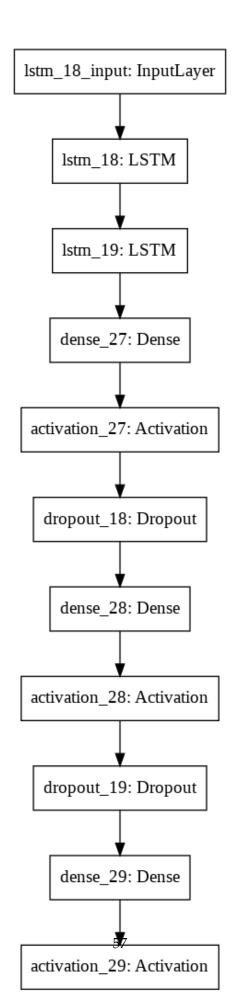
Model: "sequential_9"

Layer (type)	Output Shape	Param #
lstm_18 (LSTM)	(None, 1, 100)	44800
lstm_19 (LSTM)	(None, 100)	80400
dense_27 (Dense)	(None, 100)	10100
activation_27 (Activation)	(None, 100)	0
dropout_18 (Dropout)	(None, 100)	0
dense_28 (Dense)	(None, 100)	10100
activation_28 (Activation)	(None, 100)	0
dropout_19 (Dropout)	(None, 100)	0
dense_29 (Dense)	(None, 1)	101

activation_29 (Activation) (None, 1) 0

Total params: 145,501 Trainable params: 145,501 Non-trainable params: 0

[57]:



```
[58]: StopWatch.start("test4-train")
 history5 = train_model(model5, x_train5, y_train5, x_test5, y_test5, 100)
 StopWatch.stop("test4-train")
 Epoch 1/100
 0.3711
 Epoch 2/100
 0.3655
 Epoch 3/100
 0.3595
 Epoch 4/100
 0.3534
 Epoch 5/100
 0.3472
 Epoch 6/100
 0.3412
 Epoch 7/100
 0.3349
 Epoch 8/100
 0.3283
 Epoch 9/100
 0.3217
 Epoch 10/100
 0.3155
 Epoch 11/100
 0.3091
 Epoch 12/100
 0.3024
 Epoch 13/100
 0.2959
 Epoch 14/100
```

0.2893

```
Epoch 15/100
0.2830
Epoch 16/100
0.2770
Epoch 17/100
0.2711
Epoch 18/100
0.2655
Epoch 19/100
0.2607
Epoch 20/100
0.2560
Epoch 21/100
0.2513
Epoch 22/100
0.2468
Epoch 23/100
1/1 [============= ] - Os 92ms/step - loss: 0.1961 - val_loss:
0.2424
Epoch 24/100
0.2390
Epoch 25/100
0.2357
Epoch 26/100
0.2328
Epoch 27/100
0.2299
Epoch 28/100
0.2270
Epoch 29/100
0.2246
Epoch 30/100
0.2219
```

```
Epoch 31/100
0.2195
Epoch 32/100
0.2167
Epoch 33/100
0.2141
Epoch 34/100
0.2114
Epoch 35/100
0.2085
Epoch 36/100
0.2052
Epoch 37/100
0.2015
Epoch 38/100
0.1975
Epoch 39/100
0.1932
Epoch 40/100
0.1887
Epoch 41/100
0.1840
Epoch 42/100
0.1804
Epoch 43/100
0.1773
Epoch 44/100
0.1756
Epoch 45/100
0.1761
Epoch 46/100
0.1813
```

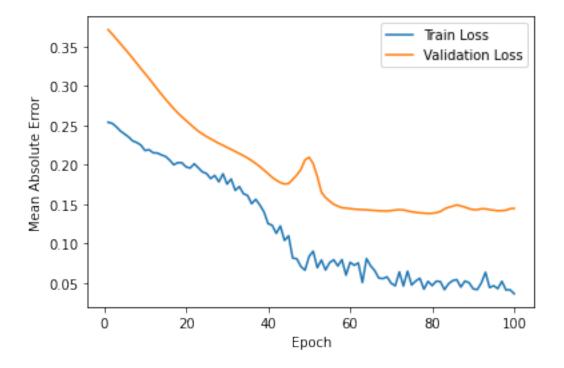
```
Epoch 47/100
0.1870
Epoch 48/100
0.1939
Epoch 49/100
0.2061
Epoch 50/100
0.2094
Epoch 51/100
0.2014
Epoch 52/100
0.1851
Epoch 53/100
0.1652
Epoch 54/100
0.1585
Epoch 55/100
0.1539
Epoch 56/100
0.1499
Epoch 57/100
0.1475
Epoch 58/100
0.1455
Epoch 59/100
0.1449
Epoch 60/100
0.1444
Epoch 61/100
0.1437
Epoch 62/100
0.1434
```

```
Epoch 63/100
0.1431
Epoch 64/100
0.1428
Epoch 65/100
0.1424
Epoch 66/100
0.1421
Epoch 67/100
Epoch 68/100
0.1414
Epoch 69/100
0.1414
Epoch 70/100
0.1418
Epoch 71/100
1/1 [============ ] - Os 83ms/step - loss: 0.0466 - val_loss:
0.1426
Epoch 72/100
0.1431
Epoch 73/100
0.1428
Epoch 74/100
0.1414
Epoch 75/100
0.1404
Epoch 76/100
0.1396
Epoch 77/100
0.1391
Epoch 78/100
0.1386
```

```
Epoch 79/100
0.1382
Epoch 80/100
0.1385
Epoch 81/100
0.1393
Epoch 82/100
0.1406
Epoch 83/100
0.1441
Epoch 84/100
0.1460
Epoch 85/100
0.1474
Epoch 86/100
0.1491
Epoch 87/100
0.1478
Epoch 88/100
0.1462
Epoch 89/100
0.1444
Epoch 90/100
0.1430
Epoch 91/100
0.1430
Epoch 92/100
0.1442
Epoch 93/100
0.1441
Epoch 94/100
0.1431
```

```
Epoch 95/100
0.1424
Epoch 96/100
1/1 [======
          =======] - Os 105ms/step - loss: 0.0430 - val_loss:
0.1417
Epoch 97/100
0.1419
Epoch 98/100
0.1423
Epoch 99/100
1/1 [============ ] - Os 98ms/step - loss: 0.0416 - val_loss:
0.1442
Epoch 100/100
0.1447
```

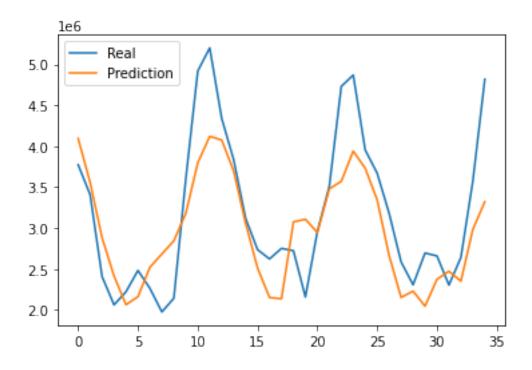
[59]: loss_plot(history5)



```
[60]: StopWatch.start("test4-predict")
predicted_model(model5, x_test5, y_test5, scaler5)
StopWatch.stop("test4-predict")
```

WARNING:tensorflow:6 out of the last 11 calls to <function
Model.make_predict_function.<locals>.predict_function at 0x7f46b062a4d0>
triggered tf.function retracing. Tracing is expensive and the excessive number
of tracings could be due to (1) creating @tf.function repeatedly in a loop, (2)
passing tensors with different shapes, (3) passing Python objects instead of
tensors. For (1), please define your @tf.function outside of the loop. For (2),
@tf.function has experimental_relax_shapes=True option that relaxes argument
shapes that can avoid unnecessary retracing. For (3), please refer to
https://www.tensorflow.org/guide/function#controlling_retracing and
https://www.tensorflow.org/api_docs/python/tf/function for more details.

The RMSE is: 587340.7243



[171]: StopWatch.benchmark()

```
| DISTRIB_DESCRIPTION | "Ubuntu 18.04.5 LTS"
| DISTRIB_ID
                  | Ubuntu
| DISTRIB_RELEASE
                  | 18.04
HOME URL
                  | "https://www.ubuntu.com/"
| ID
                  ubuntu
| ID_LIKE
                  | debian
                  | "Ubuntu"
| NAME
| PRETTY_NAME
                  | "Ubuntu 18.04.5 LTS"
| PRIVACY_POLICY_URL | "https://www.ubuntu.com/legal/terms-and-policies
/privacy-policy" |
| SUPPORT_URL
                  | "https://help.ubuntu.com/"
| UBUNTU_CODENAME
                  bionic
                  | "18.04.5 LTS (Bionic Beaver)"
| VERSION
| VERSION_CODENAME
                  | bionic
| VERSION_ID
                  | "18.04"
                  1 2
| cpu_count
| mem.active
                  | 1.4 GiB
| mem.available
                  | 11.5 GiB
mem.free
                  | 9.1 GiB
                  | 2.0 GiB
| mem.inactive
| mem.percent
                  | 9.5 %
                  | 12.7 GiB
mem.total
mem.used
                  | 1.4 GiB
| 3.7.10 (default, Feb 20 2021, 21:17:23)
| python
```

```
[GCC 7.5.0]
| python.pip
               | 19.3.1
| python.version
               | 3.7.10
| sys.platform
               | linux
               | x86_64
uname.machine
| uname.node
               | d91aa3bf059f
| uname.processor
               | x86_64
| uname.release
               4.19.112+
| uname.system
               | Linux
               | #1 SMP Thu Jul 23 08:00:38 PDT 2020
| uname.version
l user
               | collab
| Name
                  | Status
                             Time |
                                     Sum | Start
              | User | OS
                           | Version
|------
| data-load
                  ok
                          | 0.051 | 0.051 | 2021-05-02 08:10:29 |
| data-preprocess_1
               | ok
                          | 0.013 | 0.013 | 2021-05-02 08:10:31 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| data-preprocess 2
                  ok
                          0.009 | 0.076 | 2021-05-02 08:25:42 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| compile
                  lok
                          0.635 |
                                   5.718 | 2021-05-02 08:25:42 |
lok
                          | 11.979 | 97.234 | 2021-05-02 08:25:44 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
                  ok
                          | 0.954 |
                                   7.557 | 2021-05-02 08:25:56 |
| predict
| test1-data-preprocess_1 | ok
                       | 0.003 |
                                   0.017 | 2021-05-02 08:26:47 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test1-data-preprocess_2 | ok
                          | 0.016 |
                                   0.031 | 2021-05-02 08:26:47 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test2-compile
              l ok
                      | 0.629 | 3.637 | 2021-05-02 08:27:21 |
```

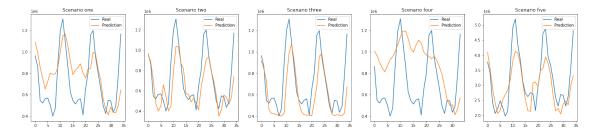
```
| test1-train
                      l ok
                                | 11.914 | 35.913 | 2021-05-02 08:26:54 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test1-predict
                      ok
                                0.945
                                           3.791 | 2021-05-02 08:27:09 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test2-data-preprocess 1 | ok
                                           0.007 | 2021-05-02 08:27:13 |
                                0.004
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test2-data-preprocess_2 | ok
                                I 0.008 I
                                           0.014 | 2021-05-02 08:27:14 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test2-train
                      lok
                                | 12.468 | 24.362 | 2021-05-02 08:27:23 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
                      ok
                                0.975
                                           1.884 | 2021-05-02 08:27:39 |
| test2-predict
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test3-data-preprocess_1 | ok
                                0.001
                                           0.006 | 2021-05-02 08:29:07 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test3-data-preprocess_2 | ok
                               0.006
                                           0.025 | 2021-05-02 08:29:08 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
                                0.602 |
                      l ok
                                           5.484 | 2021-05-02 08:31:36 |
| test3-compile
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
                                | 20.282 | 138.311 | 2021-05-02 08:31:37 |
| test3-train
                      lok
| test3-predict
                      ok
                                0.979
                                           7.888 | 2021-05-02 08:31:57 |
| test4-data-preprocess 1 | ok
                                           0.032 | 2021-05-02 08:31:58 |
                               0.019
| test4-data-preprocess_2 | ok
                               0.009 |
                                           0.016 | 2021-05-02 08:31:59 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test4-compile
                      ok
                                0.576
                                           2.416 | 2021-05-02 08:32:52 |
| test4-train
                      ok
                                | 11.858 | 47.539 | 2021-05-02 08:32:53 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
| test4-predict
                      ok
                               | 1.777 |
                                           5.596 | 2021-05-02 08:33:05 |
| d91aa3bf059f | collab | Linux | #1 SMP Thu Jul 23 08:00:38 PDT 2020 |
   ------
# csv,timer,status,time,sum,start,tag,uname.node,user,uname.system,platform.vers
# csv,data-load,ok,0.051,0.051,2021-05-02 08:10:29,,d91aa3bf059f,collab,Linux,#1
SMP Thu Jul 23 08:00:38 PDT 2020
# csv,data-preprocess_1,ok,0.013,0.013,2021-05-02
08:10:31,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,data-preprocess_2,ok,0.009,0.076,2021-05-02
08:25:42,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,compile,ok,0.635,5.718,2021-05-02 08:25:42,,d91aa3bf059f,collab,Linux,#1
SMP Thu Jul 23 08:00:38 PDT 2020
# csv,train,ok,11.979,97.234,2021-05-02 08:25:44,,d91aa3bf059f,collab,Linux,#1
SMP Thu Jul 23 08:00:38 PDT 2020
```

```
# csv,predict,ok,0.954,7.557,2021-05-02 08:25:56,,d91aa3bf059f,collab,Linux,#1
SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test1-data-preprocess_1,ok,0.003,0.017,2021-05-02
08:26:47,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test1-data-preprocess 2,ok,0.016,0.031,2021-05-02
08:26:47,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test2-compile,ok,0.629,3.637,2021-05-02
08:27:21,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test1-train,ok,11.914,35.913,2021-05-02
08:26:54,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test1-predict,ok,0.945,3.791,2021-05-02
08:27:09,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test2-data-preprocess_1,ok,0.004,0.007,2021-05-02
08:27:13, d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test2-data-preprocess_2,ok,0.008,0.014,2021-05-02
08:27:14,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test2-train,ok,12.468,24.362,2021-05-02
08:27:23,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test2-predict,ok,0.975,1.884,2021-05-02
08:27:39, d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test3-data-preprocess 1,ok,0.001,0.006,2021-05-02
08:29:07, d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test3-data-preprocess_2,ok,0.006,0.025,2021-05-02
08:29:08,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test3-compile,ok,0.602,5.484,2021-05-02
08:31:36,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test3-train,ok,20.282,138.311,2021-05-02
08:31:37,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test3-predict,ok,0.979,7.888,2021-05-02
08:31:57,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test4-data-preprocess_1,ok,0.019,0.032,2021-05-02
08:31:58,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test4-data-preprocess_2,ok,0.009,0.016,2021-05-02
08:31:59,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test4-compile,ok,0.576,2.416,2021-05-02
08:32:52,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test4-train,ok,11.858,47.539,2021-05-02
08:32:53,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
# csv,test4-predict,ok,1.777,5.596,2021-05-02
08:33:05,,d91aa3bf059f,collab,Linux,#1 SMP Thu Jul 23 08:00:38 PDT 2020
```

```
[]:
[66]: def predicted_model(model, x_test, y_test, scaler):
    y_predicted = model.predict(x_test)
    x_test = x_test.reshape(x_test.shape[0], x_test.shape[2])
    x_te_re = x_test[:,1:]
```

```
y_test = y_test.reshape(len(y_test), 1)
       inv_y_predicted = np.concatenate((y_predicted, x_te_re), axis=1)
       inv_y_predicted = scaler.inverse_transform(inv_y_predicted)[:,0]
       inv_y = np.concatenate((y_test, x_te_re), axis=1)
       inv_y = scaler.inverse_transform(inv_y)[:,0]
       rmse = np.sqrt(metrics.mean_squared_error(inv_y, inv_y_predicted))
       #print('The RMSE is: %.4f' % rmse)
      plt.plot(inv_y, label='Real')
      plt.plot(inv_y_predicted, label='Prediction')
      plt.legend()
[69]: def predicted_time_model(model, x_test, y_test, scaler, time, feature):
       y_predicted = model.predict(x_test)
       x_test = x_test.reshape(x_test.shape[0], (time*feature))
       x_te_re = x_test[:,-(feature-1):]
       y_test = y_test.reshape(len(y_test), 1)
       inv_y_predicted = np.concatenate((y_predicted, x_te_re), axis=1)
       inv_y_predicted = scaler.inverse_transform(inv_y_predicted)[:,0]
       inv_y = np.concatenate((y_test, x_te_re), axis=1)
       inv_y = scaler.inverse_transform(inv_y)[:,0]
       rmse = np.sqrt(metrics.mean_squared_error(inv_y, inv_y_predicted))
       #print('The RMSE is : %.4f' % rmse)
      plt.plot(inv_y, label='Real')
      plt.plot(inv_y_predicted, label='Prediction')
      plt.legend()
[71]: fig = plt.figure(figsize=(25,5))
     plt.subplot(151)
     plt.title('Scenario one')
     predicted_model(model, x_test, y_test, scaler)
     plt.subplot(152)
     plt.title('Scenario two')
     predicted_model(model2, x_test2, y_test2, scaler2)
     plt.subplot(153)
     plt.title('Scenario three')
     predicted_model(model3, x_test3, y_test3, scaler3)
     plt.subplot(154)
     plt.title('Scenario four')
     predicted_time_model(model4, x_test4, y_test4, scaler4, months, features)
     plt.subplot(155)
```

plt.title('Scenario five') predicted_model(model5, x_test5, y_test5, scaler5)



[]: