2/18/22, 4:09 PM Algomax_LevelC

Level C

Pick a dataset and objective

Dataset:

• Dataset of Nifty Stock prices of Indian companies. (https://www.kaggle.com/rohanrao/nifty50-stock-market-data)

Problem Statement:

• Creating an Algorithm to accept a filename as an input flag and automatically develop a Time Series Model to predict stock prices for it.

Time Period

- Time period of the year of 2008 using all previous data and for 2016 using data from 2009-2015.
 - You can start with a base algorithm choice like LSTM. However, the structure of this network should not be predefined.
 - This algorithm should automatically decide the structure of the network (no. of layers, no of neurons, etc) and hyperparameters to be used in the network
 - This algorithm should accept the name of the stock to be used as input (e.g. ICICIBANK.csv) in command line while running the program.
 - The algorithm should develop a model and save it as <stock_name>.h5 after optimizing the network configuration.
 - The algorithm should write the accuracy of the final algorithm to a file.

```
import pandas as pd  #Loading pandas for creating and adjusting dataframes
import numpy as np  # Loading Numpy for creating and adjusting arrays

import matplotlib.pyplot as plt  # Loading matplotlib for visualizations
%matplotlib inline
```

```
from sklearn.preprocessing import MinMaxScaler # Loading MinMaxScaler to scale the data
          from sklearn.model selection import GridSearchCV, KFold # Loading gridsearch CV for hyperparameter tuning using Crss Validation
          from tensorflow.keras.layers import LSTM
          from tensorflow.keras.models import Sequential
                                                             #Loading Sequential model from tensor flow to craete a LSTM model
          from tensorflow.keras.layers import LSTM
                                                             #Loading LSTM model from tensor flow to craete a ANN model
          from tensorflow.keras.layers import Dense
                                                             #Loading Dense layer from tensor flow to craete a Neural network layers
          from tensorflow.keras.layers import Dropout
                                                             #Loading Drop out Layer
          from tensorflow.keras.optimizers import Adam
                                                             #Loading Adam optimizer for LSTM
          from sklearn.metrics import mean squared error
                                                             #Loading Mean Squared Error to to check the error metrics of the model
          from sklearn.metrics import mean absolute percentage error #Loading Mean absolute percentage error to to check the error metrics
          from sklearn.metrics import r2 score
                                                     #Loading R^2 (R squared) to check the efficiency metrics of the model
In [39]:
          # Get input from user of a file name
          csv file = input("Enter CSV file name : ")
          import os
          #to get the current working directory
          directory = os.getcwd()
          print(directory)
          #Reading the file name given as input
          csv file = csv file.upper()
          file name = ("{}\\Nifty Stock prices of Indian companies\{}.csv".format(directory,csv file))
          df = pd.read csv(file name)
         Enter CSV file name : icicibank
         C:\Users\cricl\Documents\AIQ4\LEVEL C
In [40]:
          # Preprocessing the data
          def preprocessing(df, year):
              for i in '%Y-%m-%d', '%d-%m-%Y', '%m-%d-%Y':
                  df["Date"] = pd.to datetime(df.Date, format = i) #Converting the Date column to datetime fromat
                  df["Year"] = pd.to datetime(df.Date, format = i).dt.year # XCreating a Year Column in DF to split the data
              ### If year input in 2008
              # Create data between 2000 to 2008
```

```
#Considering only DATE, Closea and Year Column
if year == 2008:
    start year = year-8
    end year = year
    data = df.loc[df["Year"] <= year]</pre>
    data = data[["Date", "Close", "Year"]]
    data.set index('Date', inplace=True)
### If year input is 2016
# Create data between 2009 to 2016
#Considering only DATE, Closea and Year Column
if year == 2016:
    start year = year-7
    end year = year
    data = df[(df["Year"] <= end year) & (df["Year"] >= start year)]
    data = data[["Date","Close","Year"]]
    data.set index('Date', inplace=True)
# Create Test data and Train data as per inputs given (Input is Year)
Train data = data.loc[data["Year"] < end year]</pre>
Train data = Train data["Close"]
Train data = Train data.to frame()
Test data = data[(data["Year"] == end year)]
Test data = Test data["Close"]
Test data = Test data.to frame()
return Train data, Test data, data
```

```
In [41]:
# Creating Train and Test data with the function created
Train_data = preprocessing(df,2008)[0]
```

```
Test data = preprocessing(df,2008)[1]
          data = preprocessing(df,2008)[2]
In [42]:
          #Visualizing the Close column
          data = data.drop(["Year"], axis = 1)
          data.plot( y = "Close", figsize=(12,6))
          <AxesSubplot:xlabel='Date'>
Out[42]:
                                                                                                         Close
          1400
          1200
          1000
           800
           600
           400
           200
                                                             Date
In [43]:
          # Scaling the data using MinMaxScaler and fit the tain and test data
          scaler = MinMaxScaler(feature_range = (0,1))
          scaled array train = scaler.fit transform(Train data)
          scaled array test = scaler.fit transform(Test data)
In [44]:
          # Creating the time series based on time step with the scaled data
          ## Time step provided is 90
```

```
def create time series(scaled array,time step):
              X = []
              Y = []
              time step = time step
              for i in range(time step,len(scaled array)):
                  X.append(scaled array[i-time step:i,0])
                  Y.append(scaled array[i,0])
              X = np.array(X)
              Y = np.array(Y)
              return (X,Y)
In [45]:
          #Creating X train, Y train, X test, Y test using the above created function
          X train = create time series(scaled array train,90)[0]
          Y train = create time series(scaled array train,90)[1]
          X test = create time series(scaled array test,90)[0]
          Y test = create time series(scaled array test,90)[1]
In [46]:
          X train reshaped = np.reshape(X train,(X train.shape[0],X train.shape[1],1))
In [47]:
          X test reshaped = np.reshape(X test,(X test.shape[0],X test.shape[1],1))
In [ ]:
```

Hyperparameters all at once

- The hyperparameter optimization was carried out by taking 2 hyperparameters at once. We may have missed the best values. The performance can be further improved by finding the optimum values of hyperparameters all at once given by the code snippet below.
- This process is computationally expensive.

2/18/22, 4:09 PM Algomax_LevelC

def create_model(neuron1,neuron2,learning_rate,dropout_rate,activation_function,): model = Sequential() model.add(LSTM(neuron1, activation=activation_function,return_sequences = True, input_shape=(X_train.shape[1],1))) model.add(Dropout(0.2)) model.add(LSTM(neuron2, activation=activation_function,return_sequences = False)) model.add(Dropout(0.2)) model.add(Dense(1)) adam = Adam(learning_rate = learning_rate) model.compile(optimizer=adam, loss='mse') return model # Create the model from tensorflow.keras.wrappers.scikit_learn import KerasRegressor model = KerasRegressor(build_fn = create_model,verbose = 0,batch_size= 10,epochs = 3) # Define the grid search parameters learning_rate = [0.001,0.01,0.1] dropout_rate = [0.01,0.1,0.2] activation_function = ['softmax','relu','tanh','linear'] neuron1 = [50,60,70] neuron2 = [50,60,70] # Make a dictionary of the grid search parameters param_grids = dict(neuron1=neuron1, neuron2=neuron2, activation_function=activation_function, learning_rate = learning_rate, dropout_rate = dropout_rate) # Build and fit the GridSearchCV grid = GridSearchCV(estimator = model,param_grid = param_grids,cv = KFold(),verbose = 10) grid_result = grid.fit(X_train_reshaped,Y_train) # Summarize the results print('Best : {}, using {}'.format(grid_result.best_params_)) means = grid_result.cv_results_['mean_test_score'] stds = grid_result.cv_results_['std_test_score'] params = grid_result.cv_results_['params'] for mean, stdev, param in zip(means, stds, params): print('{},{}) with: {}'.format(mean, stdev, params_[activation_function] learning_rate = grid_result.best_params_[learning_rate] activation_function = grid_result.best_params_[neuron1] neuron2 = grid_result.best_params_[neuron2] dropout_rate,learning_rate,activation_function,neuron1,neuron2)

```
In [48]:
          # Creating the function to optimize the Hyper parameters
          # In this case Learning rate and dropout rate
          def optimize params(X,Y):
              #creating model to use in the grid serach CV
              def create model(learning rate, dropout rate):
                   model = Sequential()
                  model.add(LSTM(50, activation="relu",return sequences =True, input shape=(X train.shape[1],1)))
                  model.add(Dropout(0.2))
                  model.add(LSTM(50, activation="relu", return sequences =False))
                  model.add(Dropout(0.2))
                  model.add(Dense(1))
                   adam = Adam(learning rate = learning rate)
                  model.compile(optimizer=adam, loss='mse')
                   return model
              # Create the model
              from tensorflow.keras.wrappers.scikit learn import KerasRegressor
              model = KerasRegressor(build fn = create model, verbose = 0, batch size= 10, epochs = 3)
              # Define the grid search parameters
              learning rate = [0.001, 0.01, 0.1]
              dropout rate = [0.01, 0.1, 0.2]
              # Make a dictionary of the grid search parameters
```

In [49]:

```
# Using the function optimize_params to check for optimum values
Best_params= optimize_params(X_train_reshaped,Y_train)
```

```
Fitting 5 folds for each of 9 candidates, totalling 45 fits
[CV 1/5; 1/9] START dropout rate=0.01, learning rate=0.001......
[CV 1/5; 1/9] END dropout rate=0.01, learning rate=0.001;, score=-0.000 total time= 43.3s
[CV 2/5; 1/9] START dropout rate=0.01, learning rate=0.001......
[CV 2/5; 1/9] END dropout rate=0.01, learning rate=0.001;, score=-0.000 total time= 41.6s
[CV 3/5; 1/9] START dropout rate=0.01, learning rate=0.001......
[CV 3/5; 1/9] END dropout rate=0.01, learning rate=0.001;, score=-0.000 total time= 46.4s
[CV 4/5; 1/9] START dropout rate=0.01, learning rate=0.001......
[CV 4/5; 1/9] END dropout rate=0.01, learning rate=0.001;, score=-0.001 total time= 41.2s
[CV 5/5; 1/9] START dropout rate=0.01, learning rate=0.001......
[CV 5/5; 1/9] END dropout rate=0.01, learning rate=0.001;, score=-0.002 total time= 44.6s
[CV 1/5; 2/9] START dropout rate=0.01, learning rate=0.01......
[CV 1/5; 2/9] END dropout rate=0.01, learning rate=0.01;, score=-0.000 total time= 39.1s
[CV 2/5; 2/9] START dropout rate=0.01, learning rate=0.01......
[CV 2/5: 2/9] END dropout rate=0.01, learning rate=0.01;, score=-0.001 total time= 38.7s
[CV 3/5; 2/9] START dropout rate=0.01, learning rate=0.01......
[CV 3/5; 2/9] END dropout rate=0.01, learning rate=0.01;, score=-0.004 total time= 36.8s
[CV 4/5; 2/9] START dropout rate=0.01, learning rate=0.01......
[CV 4/5; 2/9] END dropout_rate=0.01, learning_rate=0.01;, score=nan total time= 46.1s
```

[CV 5/5; 2/9] START dropout rate=0.01, learning rate=0.01		
[CV 5/5; 2/9] END dropout_rate=0.01, learning_rate=0.01;, score=-0.019 total time= 42.6s		
[CV 1/5; 3/9] START dropout_rate=0.01, learning_rate=0.1		
[CV 1/5; 3/9] END dropout_rate=0.01, learning_rate=0.1;, score=nan total time= 47.3s		
[CV 2/5; 3/9] START dropout_rate=0.01, learning_rate=0.1		
[CV 2/5; 3/9] END dropout_rate=0.01, learning_rate=0.1;, score=nan total time= 46.4s		
[CV 3/5; 3/9] START dropout_rate=0.01, learning_rate=0.1		
[CV 3/5; 3/9] END dropout_rate=0.01, learning_rate=0.1;, score=nan total time= 44.9s		
[CV 4/5; 3/9] START dropout_rate=0.01, learning_rate=0.1		
[CV 4/5; 3/9] END dropout_rate=0.01, learning_rate=0.1;, score=nan total time= 42.7s		
[CV 5/5; 3/9] START dropout_rate=0.01, learning_rate=0.1		
[CV 5/5; 3/9] END dropout_rate=0.01, learning_rate=0.1;, score=nan total time= 32.3s		
[CV 1/5; 4/9] START dropout_rate=0.1, learning_rate=0.001		
[CV 1/5; 4/9] END dropout_rate=0.1, learning_rate=0.001;, score=-0.000 total time= 47.1s		
[CV 2/5; 4/9] START dropout_rate=0.1, learning_rate=0.001		
[CV 2/5; 4/9] END dropout_rate=0.1, learning_rate=0.001;, score=-0.000 total time= 40.7s		
[CV 3/5; 4/9] START dropout_rate=0.1, learning_rate=0.001		
[CV 3/5; 4/9] END dropout_rate=0.1, learning_rate=0.001;, score=-0.000 total time= 39.8s		
[CV 4/5; 4/9] START dropout_rate=0.1, learning_rate=0.001		
[CV 4/5; 4/9] END dropout_rate=0.1, learning_rate=0.001;, score=-0.001 total time= 41.6s		
[CV 5/5; 4/9] START dropout_rate=0.1, learning_rate=0.001		
[CV 5/5; 4/9] END dropout_rate=0.1, learning_rate=0.001;, score=-0.003 total time= 44.5s		
[CV 1/5; 5/9] START dropout_rate=0.1, learning_rate=0.01		
[CV 1/5; 5/9] END dropout_rate=0.1, learning_rate=0.01;, score=-0.010 total time= 44.0s		
[CV 2/5; 5/9] START dropout_rate=0.1, learning_rate=0.01		
[CV 2/5; 5/9] END dropout_rate=0.1, learning_rate=0.01;, score=-0.000 total time= 47.2s		
[CV 3/5; 5/9] START dropout_rate=0.1, learning_rate=0.01		
[CV 3/5; 5/9] END dropout_rate=0.1, learning_rate=0.01;, score=-0.001 total time= 36.7s		
[CV 4/5; 5/9] START dropout_rate=0.1, learning_rate=0.01		
[CV 4/5; 5/9] END dropout_rate=0.1, learning_rate=0.01;, score=-0.000 total time= 47.3s		
[CV 5/5; 5/9] START dropout_rate=0.1, learning_rate=0.01		
[CV 5/5; 5/9] END dropout rate=0.1, learning rate=0.01;, score=-0.068 total time= 38.6s		
[CV 1/5; 6/9] START dropout_rate=0.1, learning_rate=0.1		
[CV 1/5; 6/9] END dropout_rate=0.1, learning_rate=0.1;, score=nan total time= 43.0s		
[CV 2/5; 6/9] START dropout_rate=0.1, learning_rate=0.1		
[CV 2/5; 6/9] END dropout_rate=0.1, learning_rate=0.1;, score=nan total time= 46.3s		
[CV 3/5; 6/9] START dropout_rate=0.1, learning_rate=0.1		
[CV 3/5; 6/9] END dropout_rate=0.1, learning_rate=0.1;, score=nan total time= 41.1s		
[CV 4/5; 6/9] START dropout_rate=0.1, learning_rate=0.1		
[CV 4/5; 6/9] END dropout_rate=0.1, learning_rate=0.1;, score=nan total time= 46.1s		
[CV 5/5; 6/9] START dropout_rate=0.1, learning_rate=0.1		
[CV 5/5; 6/9] END dropout_rate=0.1, learning_rate=0.1;, score=nan total time= 46.1s		
[CV 1/5; 7/9] START dropout_rate=0.2, learning_rate=0.001		
[CV 1/5; 7/9] END dropout_rate=0.2, learning_rate=0.001;, score=-0.000 total time= 51.7s		
[0. 1/5, //5] the dispose_idec=0.2, fedining_idec=0.001,, score= 0.000 total time= 51./5		

2/18/22, 4:09 PM

```
[CV 2/5; 7/9] START dropout rate=0.2, learning rate=0.001......
[CV 2/5; 7/9] END dropout rate=0.2, learning rate=0.001;, score=-0.001 total time= 48.4s
[CV 3/5; 7/9] START dropout rate=0.2, learning rate=0.001......
[CV 3/5; 7/9] END dropout rate=0.2, learning rate=0.001;, score=-0.000 total time= 41.9s
[CV 4/5; 7/9] START dropout rate=0.2, learning rate=0.001......
[CV 4/5; 7/9] END dropout rate=0.2, learning rate=0.001;, score=-0.001 total time= 47.0s
[CV 5/5; 7/9] START dropout rate=0.2, learning rate=0.001......
[CV 5/5; 7/9] END dropout rate=0.2, learning rate=0.001;, score=-0.002 total time= 34.7s
[CV 1/5; 8/9] START dropout rate=0.2, learning rate=0.01.....
[CV 1/5; 8/9] END dropout rate=0.2, learning rate=0.01;, score=-0.037 total time= 34.2s
[CV 2/5; 8/9] START dropout rate=0.2, learning rate=0.01.....
[CV 2/5; 8/9] END dropout rate=0.2, learning rate=0.01;, score=-0.000 total time= 32.8s
[CV 3/5; 8/9] START dropout rate=0.2, learning rate=0.01.....
[CV 3/5; 8/9] END dropout rate=0.2, learning rate=0.01;, score=-0.001 total time= 33.5s
[CV 4/5; 8/9] START dropout rate=0.2, learning rate=0.01.....
[CV 4/5; 8/9] END dropout rate=0.2, learning rate=0.01;, score=-0.001 total time= 33.1s
[CV 5/5; 8/9] START dropout rate=0.2, learning rate=0.01.....
[CV 5/5; 8/9] END dropout rate=0.2, learning rate=0.01;, score=-0.021 total time= 36.3s
[CV 1/5; 9/9] START dropout rate=0.2, learning rate=0.1......
[CV 1/5; 9/9] END dropout rate=0.2, learning rate=0.1;, score=nan total time= 21.0s
[CV 2/5; 9/9] START dropout rate=0.2, learning rate=0.1......
[CV 2/5; 9/9] END dropout rate=0.2, learning rate=0.1;, score=nan total time= 22.2s
[CV 3/5; 9/9] START dropout rate=0.2, learning rate=0.1......
[CV 3/5; 9/9] END dropout rate=0.2, learning rate=0.1;, score=nan total time= 22.4s
[CV 4/5; 9/9] START dropout rate=0.2, learning rate=0.1......
[CV 4/5; 9/9] END dropout rate=0.2, learning rate=0.1;, score=nan total time= 21.3s
[CV 5/5; 9/9] START dropout rate=0.2, learning rate=0.1......
[CV 5/5; 9/9] END dropout rate=0.2, learning rate=0.1;, score=nan total time= 21.7s
C:\Users\cricl\anaconda3\lib\site-packages\sklearn\model selection\ search.py:922: UserWarning: One or more of the test scores are
non-finite: [-0.00071926
                              nan
                                         nan -0.001048
                                                        -0.01608253
                                                                           nan
 -0.00067702 -0.01206404
                              nan]
 warnings.warn(
Best: -0.0006770218984456733, using {'dropout rate': 0.2, 'learning rate': 0.001}
-0.0007192629214841873,0.0006151524662536869 with: {'dropout rate': 0.01, 'learning rate': 0.001}
nan, nan with: {'dropout rate': 0.01, 'learning rate': 0.01}
nan,nan with: {'dropout rate': 0.01, 'learning rate': 0.1}
-0.0010480020981049165,0.0012110273238650263 with: {'dropout rate': 0.1, 'learning rate': 0.001}
-0.016082530264975504,0.026455658525045722 with: {'dropout rate': 0.1, 'learning rate': 0.01}
nan,nan with: {'dropout rate': 0.1, 'learning rate': 0.1}
-0.0006770218984456733,0.00048233702498975956 with: {'dropout rate': 0.2, 'learning rate': 0.001}
-0.012064036238007248,0.014734933005019568 with: {'dropout rate': 0.2, 'learning rate': 0.01}
nan,nan with: {'dropout rate': 0.2, 'learning rate': 0.1}
```

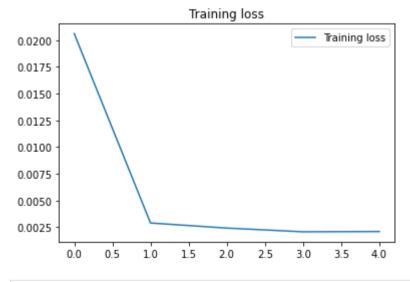
Capturing the best parameters dropout_rate = Best_params[dropout_rate] learning_rate = Best_params[learning_rate]

```
In [52]:
          Best params
         {'dropout_rate': 0.2, 'learning_rate': 0.001}
Out[52]:
In [53]:
          dropout rate = 0.2
          learning rate = 0.001
In [62]:
          ## Creating the ,model using the optimum parameters
          model LSTM = Sequential()
          model LSTM.add(LSTM(50, activation="relu", return sequences =True, input shape=(X train.shape[1],1)))
          model LSTM.add(Dropout(dropout rate))
          model LSTM.add(LSTM(50, activation="relu", return sequences =False))
          model LSTM.add(Dropout(dropout rate))
          model LSTM.add(Dense(1))
          adam = Adam(learning rate = learning rate)
          # model compiling using parameters
          model LSTM.compile(optimizer=adam, loss='mse')
          model LSTM.summary()
          #Fitting the data into the model
          hist = model LSTM.fit(X train reshaped,Y train,epochs=5,verbose = 2)
```

Model: "sequential 47"

Layer (type)	Output Shape	Param #
lstm_94 (LSTM)	(None, 90, 50)	10400
dropout_94 (Dropout)	(None, 90, 50)	0
lstm_95 (LSTM)	(None, 50)	20200
dropout_95 (Dropout)	(None, 50)	0
dense_47 (Dense)	(None, 1)	51

```
Total params: 30,651
         Trainable params: 30,651
         Non-trainable params: 0
         Epoch 1/5
          60/60 - 3s - loss: 0.0206
          Epoch 2/5
         60/60 - 3s - loss: 0.0029
          Epoch 3/5
         60/60 - 3s - loss: 0.0024
          Epoch 4/5
         60/60 - 3s - loss: 0.0021
          Epoch 5/5
         60/60 - 3s - loss: 0.0021
In [63]:
          ## Creating the H5 file and svaing the model
          model.save('{}.h5'.format(csv file))
In [64]:
          # Plotting the loss function
          plt.plot(hist.history['loss'], label='Training loss')
          plt.title("Training loss")
          plt.legend()
         <matplotlib.legend.Legend at 0x19846f3e5e0>
Out[64]:
```



```
# Predicting the train and test values
In [65]:
          train predict=model.predict(X train reshaped)
          test predict=model.predict(X test reshaped)
In [66]:
          #Inverse transforming the values to regain the unscaled values
          train predict=scaler.inverse transform(train predict)
          test predict=scaler.inverse transform(test predict)
In [67]:
          #Plotting the train and test values
          day new=np.arange(len(train predict))
          day pred=np.arange(len(train predict),len(train predict)+len(test predict))
          plt.plot(day new,train predict)
          plt.plot(day pred,test predict)
          [<matplotlib.lines.Line2D at 0x19847051f10>]
Out[67]:
          1200
          1000
           800
           600
           400
                          500
                                    1000
                                              1500
                                                         2000
In [70]:
          # Created a function to evaluate the metrics of all the models
          def Evaluate models(model):
              y_pred = model.predict(X_test_reshaped)
              MSE = mean squared error(Y test, y pred)
```

RMSE = mean_squared_error(Y_test, y_pred, squared=False)

2/18/22, 4:09 PM Algomax_LevelC

```
R2 = r2 score(Y test, y pred)
              ADJ R2 = 1 - (1-R2)*(len(Y train)-1)/(X train.shape[0]-X train.shape[1]-1)
              MAPE = mean absolute percentage error(Y test, y pred)
              return (MSE,RMSE,R2,ADJ_R2,MAPE)
In [73]:
          # Loading the metrics into DF using above function
          Metrics LSTM = Evaluate models(model LSTM)
          MSE LSTM = Metrics LSTM[0]
          RMSE LSTM = Metrics LSTM[1]
          R2 LSTM = Metrics LSTM[2]
          ADJ R2 LSTM = Metrics LSTM[3]
          MAPE LSTM = Metrics_LSTM[4]
In [74]:
          Metrics={'Models':["Neural Network"],
                 'MSE' :[MSE LSTM],
                 'RMSE' : [RMSE LSTM],
                 'R2' : [R2 LSTM],
                 'ADJ R2' : [ADJ R2 LSTM,],
                 'MAPE' : [MAPE LSTM]
          Metrics=pd.DataFrame(Metrics)
          # Saving the DF into CSV file
          Metrics.to csv("Metrics.csv")
In [75]:
          Metrics
Out[75]:
                  Models
                             MSE
                                     RMSE
                                                R2 ADJ R2
                                                                  MAPE
          0 Neural Network 0.004493 0.067029 0.780818 0.770015 3.496228e+12
 In [ ]:
```