

```
[3]: # Importing libraries to be used
import numpy as np # for linear algebra
import pandas as pd # data preprocessing
import matplotlib.pyplot as plt # data visualization library
import seaborn as sns # data visualization library
%matplotlib inline

# Ignoring warnings
warnings.filterwarnings('ignore') # ignore warnings

# from sklearn.preprocessing import MinMaxScaler # for normalization
from keras.models import Sequential
from keras.layers import Dense, Dropout, LSTM, Bidirectional

In [5]: df = pd.read_csv('GOOG.csv') # data importing
df.head(10) # fetching first 10 rows of dataset

Out [5]:
```

	Symbol	date	close	high	low	open	volume	adjClose	adjHigh	adjLow	adjOpen	adjVolume	divCash	splitFactor
0	GOOG	2016-06-14 00:00:00+00:00	718.27	722.47	713.1200	716.48	1300605	718.27	722.47	713.1200	716.48	1300605	0.0	1.0
1	GOOG	2016-06-15 00:00:00+00:00	718.92	722.98	713.1200	716.48	1214517	718.92	722.98	713.1200	716.48	1214517	0.0	1.0
2	GOOG	2016-06-16 00:00:00+00:00	710.36	716.65	703.2600	714.91	1982471	710.36	716.65	703.2600	714.91	1982471	0.0	1.0
3	GOOG	2016-06-17 00:00:00+00:00	691.72	708.82	688.4515	698.46	2002538	693.71	702.48	688.4515	708.65	3402357	0.0	1.0
4	GOOG	2016-06-20 00:00:00+00:00	695.94	702.48	693.4105	698.77	1405634	695.94	702.77	692.0100	698.40	2002538	0.0	1.0
5	GOOG	2016-06-21 00:00:00+00:00	695.94	702.77	692.0100	698.40	1405634	695.94	702.77	692.0100	698.40	1405634	0.0	1.0
6	GOOG	2016-06-22 00:00:00+00:00	697.46	700.86	693.0819	699.46	1184318	697.46	700.86	693.0819	699.46	1184318	0.0	1.0
7	GOOG	2016-06-23 00:00:00+00:00	701.87	701.95	687.0000	697.45	2174145	701.87	701.95	687.0000	697.45	2174145	0.0	1.0
8	GOOG	2016-06-24 00:00:00+00:00	675.22	689.40	673.4500	675.17	4449022	675.22	689.40	673.4500	675.17	4449022	0.0	1.0
9	GOOG	2016-06-27 00:00:00+00:00	668.26	672.30	663.2840	671.00	2641085	668.26	672.30	663.2840	671.00	2641085	0.0	1.0

```
In [8]: #Shape of data
print(df.shape, df.dtypes)

Shape of data: (1258, 14)

In [7]: #Statistical description of data
df.describe()

Out [7]:
```

	close	high	low	open	volume	adjClose	adjHigh	adjLow	adjOpen	adjVolume	divCash	splitFactor
count	1258.000000	1258.000000	1258.000000	1258.000000	1.258000e+03	1258.000000	1258.000000	1258.000000	1258.000000	1.258000e+03	1258.0	1258.0
mean	1216.317007	1227.430934	1204.176430	1215.260779	1.601590e+06	1216.317007	1227.430936	1204.176430	1215.260779	1.601590e+06	0.0	1.0
std	383.33358	387.570872	378.777004	382.468995	6.960172e+05	383.33358	387.570873	378.777099	382.468995	6.960172e+05	0.0	0.0
min	668.260000	667.300000	663.284000	671.000000	3.460173e+05	668.260000	672.300000	663.284000	671.000000	3.460173e+05	0.0	1.0
25%	960.802500	968.175000	952.182500	969.050000	1.173520e+06	960.802500	968.175000	952.182500	969.050000	1.173520e+06	0.0	1.0
50%	1132.400000	1143.835000	1117.915000	1131.150000	1.412588e+06	1132.400000	1143.835000	1117.915000	1131.150000	1.412588e+06	0.0	1.0
75%	1360.565000	1374.345000	1348.557500	1361.075000	1.812156e+06	1360.565000	1374.345000	1348.557500	1361.075000	1.812156e+06	0.0	1.0
max	2521.000000	2526.990000	2498.290000	2524.920000	6.207027e+06	2521.000000	2526.990000	2498.290000	2524.920000	6.207027e+06	0.0	1.0

```
In [8]: # summary of data
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1258 entries, 0 to 1257
Data columns (total 14 columns):
 #   column      Non-Null Count  Dtype
---  --
 0   date        1258 non-null      object
 1   close       1258 non-null      float64
 2   high        1258 non-null      float64
 3   low         1258 non-null      float64
 4   open        1258 non-null      float64
 5   volume      1258 non-null      int64
 6   adjClose    1258 non-null      float64
 7   adjHigh     1258 non-null      float64
 8   adjLow      1258 non-null      float64
 9   adjOpen     1258 non-null      float64
10  adjVolume   1258 non-null      float64
11  divCash     1258 non-null      float64
12  splitFactor 1258 non-null      float64
13  dtypes      1258 non-null      object
14  memory usage: 137.7+ KB

In [9]: # checking null values
df.isnull().sum()

Out [9]:
```

	Symbol	date	close	high	low	open	volume	adjClose	adjHigh	adjLow	adjOpen	adjVolume	divCash	splitFactor
Symbol	0	0	0	0	0	0	0	0	0	0	0	0	0	0
date	0	0	0	0	0	0	0	0	0	0	0	0	0	0
close	0	0	0	0	0	0	0	0	0	0	0	0	0	0
high	0	0	0	0	0	0	0	0	0	0	0	0	0	0
low	0	0	0	0	0	0	0	0	0	0	0	0	0	0
open	0	0	0	0	0	0	0	0	0	0	0	0	0	0
volume	0	0	0	0	0	0	0	0	0	0	0	0	0	0
adjClose	0	0	0	0	0	0	0	0	0	0	0	0	0	0
adjHigh	0	0	0	0	0	0	0	0	0	0	0	0	0	0
adjLow	0	0	0	0	0	0	0	0	0	0	0	0	0	0
adjOpen	0	0	0	0	0	0	0	0	0	0	0	0	0	0
adjVolume	0	0	0	0	0	0	0	0	0	0	0	0	0	0
divCash	0	0	0	0	0	0	0	0	0	0	0	0	0	0
splitFactor	0	0	0	0	0	0	0	0	0	0	0	0	0	0
dtypes	1258	1258	1258	1258	1258	1258	1258	1258	1258	1258	1258	1258	1258	1258

```
In [10]: # df[['date', 'open', 'close']] # Extracting required columns
df[['date', 'open', 'close']] = df[['date', 'open', 'close']].apply(lambda x: x.split()[0]).astype('object')
df['date'] = df['date'].dropna(inplace=True) # Setting date column as index
df.head(10)

Out [10]:
```

	date	open	close
0	2016-06-14	716.48	718.27
1	2016-06-15	719.90	718.92
2	2016-06-16	714.91	710.36
3	2016-06-17	708.65	691.72
4	2016-06-20	698.77	693.71
5	2016-06-21	698.40	695.94
6	2016-06-22	699.05	697.46
7	2016-06-23	697.45	701.87
8	2016-06-24	675.17	675.22
9	2016-06-27	671.00	668.26

```
In [11]: # plotting open and closing price on date index
fig, ax = plt.subplots(1, 2, figsize=(20, 7))
ax[0].plot(df['open'], label='open', color='green')
ax[0].set_xlabel('Date', size=15)
ax[0].set_ylabel('Price', size=15)
ax[0].legend()

ax[1].plot(df['close'], label='close', color='red')
ax[1].set_xlabel('Date', size=15)
ax[1].set_ylabel('Price', size=15)
ax[1].legend()

fig.show()
```

```
In [12]: # normalizing all the values of all columns using MinMaxScaler
MMS = MinMaxScaler()
df.fit_transform(df)

Out [12]:
```

	date	open	close
0	2016-06-14	0.024532	0.026984
1	2016-06-15	0.025891	0.025154
2	2016-06-16	0.023605	0.020271
3	2016-06-17	0.020208	0.016581
4	2016-06-20	0.014979	0.013732
5	2016-06-21	0.014739	0.014935
6	2016-06-22	0.015135	0.015755
7	2016-06-23	0.014267	0.018135
8	2016-06-24	0.002490	0.003755
9	2016-06-27	0.000000	0.000000

```
In [13]: # splitting the data into training and test set
train_size = round(len(df) * 0.75) # Selecting 75 % for training and 25 % for testing
df.head(10)

Out [13]:
```

	date	open	close
0	2016-06-14	716.48	718.27
1	2016-06-15	719.90	718.92
2	2016-06-16	714.91	710.36
3	2016-06-17	708.65	691.72
4	2016-06-20	698.77	693.71
5	2016-06-21	698.40	695.94
6	2016-06-22	699.05	697.46
7	2016-06-23	697.45	701.87
8	2016-06-24	675.17	675.22
9	2016-06-27	671.00	668.26

```
In [14]: train_data = df[:train_size]
test_data = df[train_size:]

train_data.shape, test_data.shape

Out [14]:
```

	date	open	close
0	2016-06-14	716.48	718.27
1	2016-06-15	719.90	718.92
2	2016-06-16	714.91	710.36
3	2016-06-17	708.65	691.72
4	2016-06-20	698.77	693.71
5	2016-06-21	698.40	695.94
6	2016-06-22	699.05	697.46
7	2016-06-23	697.45	701.87
8	2016-06-24	675.17	675.22
9	2016-06-27	671.00	668.26

```
In [15]: # Function to create sequence of data for training and testing
def create_sequence(dataset):
    sequences = []
    labels = []

    for start_idx in range(50, len(dataset)): # Selecting 50 rows at a time
        sequences.append(dataset.iloc[start_idx:stop_idx])
        labels.append(dataset.iloc[stop_idx])
        start_idx += 1

    return (np.array(sequences), np.array(labels))

In [16]: train_seq, train_label = create_sequence(train_data)
train_seq.shape, test_seq.shape, train_label.shape

Out [16]:
```

	date	open	close
0	2016-06-14	716.48	718.27
1	2016-06-15	719.90	718.92
2	2016-06-16	714.91	710.36
3	2016-06-17	708.65	691.72
4	2016-06-20	698.77	693.71
5	2016-06-21	698.40	695.94
6	2016-06-22	699.05	697.46
7	2016-06-23	697.45	701.87
8	2016-06-24	675.17	675.22
9	2016-06-27	671.00	668.26

```
In [17]: # Importing Sequential from keras.models
model = Sequential()
# adding layers: dropout, LSTM, Bidirectional from keras.layers
model.add(LSTM(units=50, return_sequences=True, input_shape = (train_seq.shape[1], train_seq.shape[2])))
model.add(Dropout(0.1))
model.add(LSTM(units=50))
model.add(Dense(2))

model.compile(loss='mean_squared_error', optimizer='adam', metrics=['mean_absolute_error'])

model.summary()
```

Layer (type)	Output Shape	Param #
LSTM (LSTM)	(None, 50, 50)	169000
Dropout (Dropout)	(None, 50, 50)	0
LSTM_1 (LSTM)	(None, 50)	262000
dense (Dense)	(None, 2)	182

Total params: 39902 (128.71 KB)
Trainable params: 39892 (128.71 KB)
Non-trainable params: 0 (0.00 Byte)

```
In [18]: # Fitting the model by iterating the dataset over 100 times(100 epochs)
model.fit(train_seq, train_label, epochs=100, validation_data=(test_seq, test_label), verbose=1)

Epoch 1/100
28/28 [=====] - 22s 166ms/step - loss: 0.8097 - mean_absolute_error: 0.0685 - val_loss: 0.8255 - val_mean_absolute_error: 0.1371
Epoch 2/100
28/28 [=====] - 2s 82ms/step - loss: 0.8686e-04 - mean_absolute_error: 0.0239 - val_loss: 0.0654 - val_mean_absolute_error: 0.0054
Epoch 3/100
28/28 [=====] - 2s 83ms/step - loss: 4.7381e-04 - mean_absolute_error: 0.0162 - val_loss: 0.0641 - val_mean_absolute_error: 0.0058
Epoch 4/100
28/28 [=====] - 2s 83ms/step - loss: 4.5095e-04 - mean_absolute_error: 0.0155 - val_loss: 0.0655 - val_mean_absolute_error: 0.0058
Epoch 5/100
28/28 [=====] - 2s 80ms/step - loss: 4.4746e-04 - mean_absolute_error: 0.0154 - val_loss: 0.0645 - val_mean_absolute_error: 0.0055
Epoch 6/100
28/28 [=====] - 2s 79ms/step - loss: 4.3762e-04 - mean_absolute_error: 0.0153 - val_loss: 0.0636 - val_mean_absolute_error: 0.0058
Epoch 7/100
28/28 [=====] - 2s 81ms/step - loss: 4.0955e-04 - mean_absolute_error: 0.0148 - val_loss: 0.0667 - val_mean_absolute_error: 0.0068
Epoch 8/100
28/28 [=====] - 2s 80ms/step - loss: 4.5618e-04 - mean_absolute_error: 0.0155 - val_loss: 0.0647 - val_mean_absolute_error: 0.0067
Epoch 9/100
28/28 [=====] - 2s 82ms/step - loss: 4.0863e-04 - mean_absolute_error: 0.0147 - val_loss: 0.0646 - val_mean_absolute_error: 0.0069
Epoch 10/100
28/28 [=====] - 2s 83ms/step - loss: 3.7810e-04 - mean_absolute_error: 0.0144 - val_loss: 0.0644 - val_mean_absolute_error: 0.0066
Epoch 11/100
28/28 [=====] - 2s 81ms/step - loss: 3.9261e-04 - mean_absolute_error: 0.0141 - val_loss: 0.0640 - val_mean_absolute_error: 0.0064
Epoch 12/100
28/28 [=====] - 2s 80ms/step - loss: 4.0585e-04 - mean_absolute_error: 0.0146 - val_loss: 0.0656 - val_mean_absolute_error: 0.0068
Epoch 13/100
28/28 [=====] - 2s 83ms/step - loss: 4.0111e-04 - mean_absolute_error: 0.0145 - val_loss: 0.0650 - val_mean_absolute_error: 0.0068
Epoch 14/100
28/28 [=====] - 2s 82ms/step - loss: 3.7329e-04 - mean_absolute_error: 0.0141 - val_loss: 0.0671 - val_mean_absolute_error: 0.0071
Epoch 15/100
28/28 [=====] - 2s 82ms/step - loss: 3.6214e-04 - mean_absolute_error: 0.0139 - val_loss: 0.0646 - val_mean_absolute_error: 0.0066
Epoch 16/100
28/28 [=====] - 2s 82ms/step - loss: 3.7578e-04 - mean_absolute_error: 0.0140 - val_loss: 0.0656 - val_mean_absolute_error: 0.0068
Epoch 17/100
28/28 [=====] - 2s 83ms/step - loss: 3.5818e-04 - mean_absolute_error: 0.0137 - val_loss: 0.0637 - val_mean_absolute_error: 0.0068
Epoch 18/100
28/28 [=====] - 2s 86ms/step - loss: 3.4563e-04 - mean_absolute_error: 0.0137 - val_loss: 0.0629 - val_mean_absolute_error: 0.0069
Epoch 19/100
28/28 [=====] - 2s 82ms/step - loss: 3.3414e-04 - mean_absolute_error: 0.0134 - val_loss: 0.0651 - val_mean_absolute_error: 0.0064
Epoch 20/100
28/28 [=====] - 2s 80ms/step - loss: 3.1890e-04 - mean_absolute_error: 0.0132 - val_loss: 0.0635 - val_mean_absolute_error: 0.0065
Epoch 21/100
28/28 [=====] - 2s 81ms/step - loss: 3.0953e-04 - mean_absolute_error: 0.0130 - val_loss: 0.0647 - val_mean_absolute_error: 0.0064
Epoch 22/100
28/28 [=====] - 2s 83ms/step - loss: 3.1646e-04 - mean_absolute_error: 0.0130 - val_loss: 0.0646 - val_mean_absolute_error: 0.0064
Epoch 23/100
28/28 [=====] - 2s 81ms/step - loss: 3.1890e-04 - mean_absolute_error: 0.0126 - val_loss: 0.0619 - val_mean_absolute_error: 0.0064
Epoch 24/100
28/28 [=====] - 2s 83ms/step - loss: 3.1056e-04 - mean_absolute_error: 0.0130 - val_loss: 0.0643 - val_mean_absolute_error: 0.0064
Epoch 25/100
28/28 [=====] - 2s 82ms/step - loss: 2.9578e-04 - mean_absolute_error: 0.0124 - val_loss: 0.0667 - val_mean_absolute_error: 0.0067
Epoch 26/100
28/28 [=====] - 2s 84ms/step - loss: 2.8085e-04 - mean_absolute_error: 0.0123 - val_loss: 0.0640 - val_mean_absolute_error: 0.0064
Epoch 27/100
28/28 [=====] - 2s 80ms/step - loss: 2.8881e-04 - mean_absolute_error: 0.0125 - val_loss: 0.0657 - val_mean_absolute_error: 0.0064
Epoch 28/100
28/28 [=====] - 2s 79ms/step - loss: 2.8048e-04 - mean_absolute_error: 0.0124 - val_loss: 0.0646 - val_mean_absolute_error: 0.0064
Epoch 29/100
28/28 [=====] - 2s 82ms/step - loss: 2.9411e-04 - mean_absolute_error: 0.0127 - val_loss: 0.0638 - val_mean_absolute_error: 0.0064
Epoch 30/100
28/28 [=====] - 2s 81ms/step - loss: 2.8119e-04 - mean_absolute_error: 0.0125 - val_loss: 0.0644 - val_mean_absolute_error: 0.0064
Epoch 31/100
28/28 [=====] - 2s 81ms/step - loss: 2.9683e-04 - mean_absolute_error: 0.0127 - val_loss: 0.0648 - val_mean_absolute_error: 0.0064
Epoch 32/100
28/28 [=====] - 2s 81ms/step - loss: 2.7314e-04 - mean_absolute_error: 0.0123 - val_loss: 0.0655 - val_mean_absolute_error: 0.0064
Epoch 33/100
28/28 [=====] - 2s 80ms/step - loss: 2.8152e-04 - mean_absolute_error: 0.0124 - val_loss: 0.0643 - val_mean_absolute_error: 0.0064
Epoch 34/100
28/28 [=====] - 2s 82ms/step - loss: 2.5571e-04 - mean_absolute_error: 0.0116 - val_loss: 0.0637 - val_mean_absolute_error: 0.0064
Epoch 35/100
28/28 [=====] - 2s 80ms/step - loss: 2.1474e-04 - mean_absolute_error: 0.0120 - val_loss: 0.0619 - val_mean_absolute_error: 0.0064
Epoch 36/100
28/28 [=====] - 2s 82ms/step - loss: 2.5686e-04 - mean_absolute_error: 0.0120 - val_loss: 0.0651 - val_mean_absolute_error: 0.0064
Epoch 37/100
28/28 [=====] - 2s 79ms/step - loss: 2.4937e-04 - mean_absolute_error: 0.0117 - val_loss: 0.0639 - val_mean_absolute_error: 0.0064
Epoch 38/100
28/28 [=====] - 2s 78ms/step - loss: 2.3182e-04 - mean_absolute_error: 0.0112 - val_loss: 0.0647 - val_mean_absolute_error: 0.0064
Epoch 39/100
28/28 [=====] - 2s 79ms/step - loss: 2.4486e-04 - mean_absolute_error: 0.0114 - val_loss: 0.0663 - val_mean_absolute_error: 0.0063
Epoch 40/100
28/28 [=====] - 2s 81ms/step - loss: 2.4822e-04 - mean_absolute_error: 0.0116 - val_loss: 0.0630 - val_mean_absolute_error: 0.0064
Epoch 41/100
28/28 [=====] - 2s 87ms/step - loss:
```