Project - Stock Price Prediction

In [3]:
```
# Importing libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
from sklearn import metrics
from keras.models import Sequential
from keras.layers import Dense,LSTM,Dropout
```

In [4]:
```
# Importing data
df = pd.read_csv('1729258-1613615-Stock_Price_data_set_(1).csv')
df.head()
```

Out[4]:

	Date	Open	High	Low	Close	Adj Close	Volume
0	2018-02-05	262.000000	267.899994	250.029999	254.259995	254.259995	11896100
1	2018-02-06	247.699997	266.700012	245.000000	265.720001	265.720001	12595800
2	2018-02-07	266.579987	272.450012	264.329987	264.559998	264.559998	8981500
3	2018-02-08	267.079987	267.619995	250.000000	250.100006	250.100006	9306700
4	2018-02-09	253.850006	255.800003	236.110001	249.470001	249.470001	16906900

In [5]:
```
# Check shape of the dataset
df.shape
```

Out[5]: (1009, 7)

In [6]:
```
# Info of the dataset
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1009 entries, 0 to 1008
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   Date       1009 non-null   object 
 1   Open       1009 non-null   float64
 2   High       1009 non-null   float64
 3   Low        1009 non-null   float64
 4   Close      1009 non-null   float64
 5   Adj Close  1009 non-null   float64
 6   Volume     1009 non-null   int64  
dtypes: float64(5), int64(1), object(1)
memory usage: 55.3+ KB
```

In [7]:
```
# Description of the dataset
df.describe()
```

Out[7]:

	Open	High	Low	Close	Adj Close	Volume
count	1009.000000	1009.000000	1009.000000	1009.000000	1009.000000	1.009000e+03
mean	419.059673	425.320703	412.374044	419.000733	419.000733	7.570685e+06
std	108.537532	109.262960	107.555867	108.289999	108.289999	5.465535e+06
min	233.919998	250.649994	231.229996	233.880005	233.880005	1.144000e+06
25%	331.489990	336.299988	326.000000	331.619995	331.619995	4.091900e+06
50%	377.769989	383.010010	370.880005	378.670013	378.670013	5.934500e+06
75%	509.130005	515.630005	502.529999	509.079987	509.079987	9.322400e+06
max	692.349976	700.989990	686.090027	691.690002	691.690002	5.890430e+07

In [8]:
```
# Sum of null values
df.isnull().sum()
```

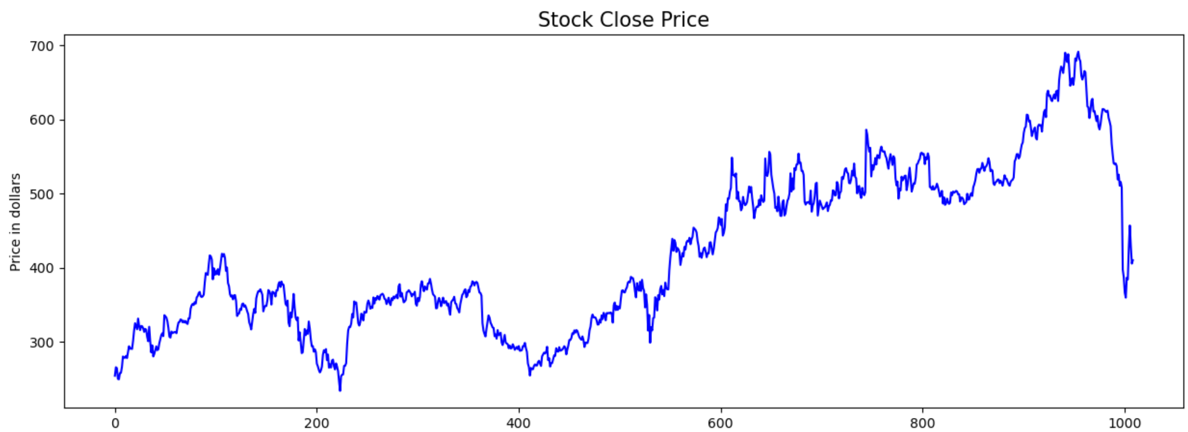
Out[8]:
```
Date         0
Open         0
High         0
Low          0
Close        0
Adj Close    0
Volume       0
dtype: int64
```

In [9]:
```
# Looking for the unique values
df.nunique()
```

Out[9]:
```
Date         1009
Open          976
High          983
Low           989
Close         988
Adj Close     988
Volume       1005
dtype: int64
```

Data Analysis

In [10]:
```
plt.figure(figsize=(15,5))
plt.plot(df['Close'], color="blue")
plt.title('Stock Close Price', fontsize=15)
plt.ylabel('Price in dollars')
plt.show()
```



In [11]:
```
# Splitting the data into training and testing sets

df_train = pd.DataFrame(df['Close'][0:int(len(df)*0.70)])          #70% use
df_test = pd.DataFrame(df['Close'][int(len(df)*0.70):int(len(df))])   #30% use

print(df_train.shape)
print(df_test.shape)
```

```
(706, 1)
(303, 1)
```

In [12]:
```
# Checking the output of training & testing sets
df_train.head()
```

Out[12]:

	Close
0	254.259995
1	265.720001
2	264.559998
3	250.100006
4	249.470001

In [13]:
```
df_test.head()
```

Out[13]:

	Close
706	476.619995
707	482.880005
708	485.000000
709	491.359985
710	490.700012

In [14]:
```
# Scaling the data
scaler = MinMaxScaler(feature_range=(0,1))
```

In [15]:
```
df_train_array = scaler.fit_transform(df_train)
df_train_array
```

Out[15]:
```
array([[0.06316048],
       [0.09867666],
       [0.09508165],
       [0.05026808],
       [0.04831561],
       [0.07459636],
       [0.07558802],
       [0.09954442],
       [0.14376913],
       [0.13834564],
       [0.13843861],
       [0.14615554],
       [0.13716804],
       [0.16131029],
       [0.18681626],
       [0.17581425],
       [0.17820065],
       [0.17513253],
       [0.2081693 ],
```

In [16]:
```
# Chekcking the shape of scaled array
df_train_array.shape
```

Out[16]: (706, 1)

In [17]:
```
# Preparing the training data

X_train = []
y_train = []

for i in range(100,df_train_array.shape[0]):
    X_train.append(df_train_array[i-100:i])
    y_train.append(df_train_array[i,0])

X_train,y_train = np.array(X_train),np.array(y_train)
```

In [18]:
```
# Building model of 4 LSTM network followed by Dropout layout

model = Sequential()

model.add(LSTM(units=50, activation = 'relu', return_sequences = True, input_sh
model.add(Dropout(0.2))

model.add(LSTM(units=60, activation = 'relu', return_sequences = True))
model.add(Dropout(0.3))

model.add(LSTM(units=80, activation = 'relu', return_sequences = True))
model.add(Dropout(0.4))

model.add(LSTM(units=120, activation = 'relu'))
model.add(Dropout(0.5))

model.add(Dense(units = 1))
```

In [19]:
```
# Checking the summary
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 100, 50)           10400

 dropout (Dropout)           (None, 100, 50)           0

 lstm_1 (LSTM)               (None, 100, 60)           26640

 dropout_1 (Dropout)         (None, 100, 60)           0

 lstm_2 (LSTM)               (None, 100, 80)           45120

 dropout_2 (Dropout)         (None, 100, 80)           0

 lstm_3 (LSTM)               (None, 120)               96480

 dropout_3 (Dropout)         (None, 120)               0
```

In [20]:
```
# Compiling & fitting the model

model.compile(optimizer = 'adam', loss = 'mean_squared_error')
hist = model.fit(X_train,y_train, epochs = 50, batch_size = 32, verbose = 2 )
```

```
Epoch 41/50
19/19 - 4s - loss: 0.0073 - 4s/epoch - 224ms/step
Epoch 42/50
19/19 - 4s - loss: 0.0065 - 4s/epoch - 225ms/step
Epoch 43/50
19/19 - 4s - loss: 0.0073 - 4s/epoch - 228ms/step
Epoch 44/50
19/19 - 4s - loss: 0.0070 - 4s/epoch - 229ms/step
Epoch 45/50
19/19 - 4s - loss: 0.0072 - 4s/epoch - 229ms/step
Epoch 46/50
19/19 - 4s - loss: 0.0075 - 4s/epoch - 227ms/step
Epoch 47/50
19/19 - 4s - loss: 0.0072 - 4s/epoch - 224ms/step
Epoch 48/50
19/19 - 4s - loss: 0.0066 - 4s/epoch - 225ms/step
Epoch 49/50
19/19 - 4s - loss: 0.0064 - 4s/epoch - 225ms/step
Epoch 50/50
19/19 - 4s - loss: 0.0072 - 4s/epoch - 223ms/step
```

In [21]:
```
df_test.head()
```

Out[21]:

	Close
706	476.619995
707	482.880005
708	485.000000
709	491.359985
710	490.700012

For prediction, we need testing data and if we look the test data from above table. We can say that we need previous days data for prediction. Hence, for prediction append the 'df_train.tail() to df_test.head()' as mentioned below:

In [22]:
```
df_train.tail()
```

Out[22]:

	Close
701	479.100006
702	480.630005
703	481.790009
704	484.670013
705	488.239990

In [23]:
```
# Append testing & training data
past_100_days = df_train.tail(100)
```

In [24]:
```
final_df = past_100_days.append(df_test, ignore_index=True)
```

In [25]:
```
# Scaling the data

input_data = scaler.fit_transform(final_df)
input_data
```

Out[25]:
```
array([[0.35299258],
       [0.40395792],
       [0.40200005],
       [0.43097681],
       [0.4459773 ],
       [0.56938454],
       [0.49941261],
       [0.4975451 ],
       [0.49266545],
       [0.5051056 ],
       [0.40148794],
       [0.42986233],
       [0.39278291],
       [0.39193951],
       [0.35507087],
       [0.36371579],
       [0.40950024],
       [0.38799362],
       [0.3758547 ],
```

In [26]:
```
# Checking shape of the input_data
input_data.shape
```

Out[26]: (403, 1)

In [27]:
```
# Preparing the testing data
X_test = []
y_test = []

for i in range(100,input_data.shape[0]):
    X_test.append(input_data[i-100:i])
    y_test.append(input_data[i,0])

X_test,y_test = np.array(X_test), np.array(y_test)
print(X_test.shape)
print(y_test.shape)
```

```
(303, 100, 1)
(303,)
```

In [28]:
```
# Making Predictions

y_pred = model.predict(X_test)
print(y_pred.shape)
```

```
10/10 [==============================] - 1s 70ms/step
(303, 1)
```

In [29]:
```
# Checking y_test
y_test
```

Out[29]:
```
array([0.35217924, 0.37103526, 0.37742098, 0.39657814, 0.39459021,
       0.43639862, 0.43278411, 0.41513293, 0.41751255, 0.47013471,
       0.46073667, 0.4033254 , 0.42588629, 0.43230216, 0.49013517,
       0.48218326, 0.49739453, 0.52170251, 0.52637129, 0.50968392,
       0.50492488, 0.46621878, 0.46468256, 0.48019515, 0.51558778,
       0.49667165, 0.54528743, 0.49146052, 0.48525552, 0.42407899,
       0.44938103, 0.45392929, 0.41989216, 0.40528327, 0.44606766,
       0.42519346, 0.41651858, 0.42793452, 0.68267123, 0.66309233,
       0.61890412, 0.59363241, 0.60914481, 0.49272575, 0.53887156,
       0.52016629, 0.54019691, 0.5676676 , 0.54143199, 0.57971616,
       0.57558954, 0.56694472, 0.60053014, 0.61414507, 0.59607223,
       0.59284922, 0.59513848, 0.57724637, 0.56784832, 0.54375121,
       0.52435321, 0.56161335, 0.58348133, 0.56326999, 0.5396246 ,
       0.57513783, 0.56664358, 0.48495438, 0.45661014, 0.47197207,
       0.40251206, 0.44200125, 0.43627821, 0.49206299, 0.47688187,
       0.48359888, 0.49498485, 0.49621975, 0.43703124, 0.4592909 ,
       0.49221355, 0.52829911, 0.48528567, 0.43121774, 0.44685075,
       0.46462244, 0.46293565, 0.48784592, 0.54134154, 0.54510671,
       0.5567337 , 0.56414345, 0.58700567, 0.58920447, 0.58158385,
```

In [30]:
```
# Checking y_pred
y_pred
```

Out[30]:
```
array([[0.3939268 ],
       [0.39372668],
       [0.39306623],
       [0.39229423],
       [0.39214325],
       [0.39306444],
       [0.39609826],
       [0.4016729 ],
       [0.40910065],
       [0.41733265],
       [0.4265015 ],
       [0.4363374 ],
       [0.44489366],
       [0.45101655],
       [0.4543289 ],
       [0.45625192],
       [0.45789665],
       [0.46020943],
       [0.46413487],
```

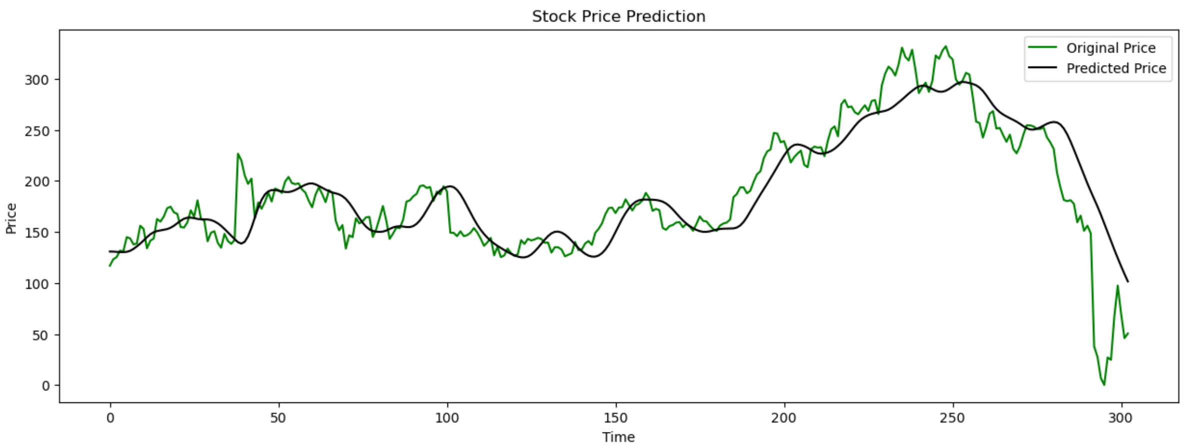
From above y_test & y_pred, we can't recognize how they are matching. hence, for that we need to scale the data.

In [31]:
```
# Scaling the data
scaler.scale_
```

Out[31]:
```
array([0.00301214])
```

In [32]:
```
scale_factor = 1/0.00301214
y_pred = y_pred * scale_factor
y_test = y_test * scale_factor
```

In [33]:
```
# Plotting graph for the result
plt.figure(figsize = (15,5))
plt.plot(y_test,'g',label = 'Original Price')
plt.plot(y_pred,'k',label = 'Predicted Price')
plt.title('Stock Price Prediction')
plt.xlabel('Time')
plt.ylabel('Price')
plt.legend()
plt.show()
```



Conclusion :

Above graph shows the relation between Actual price(Green Line) and Predicted price(Black Line) of stock for the mentioned dataset.