

BUILDING A STOCK PRICE PREDICTION MODEL: DATA COLLECTION AND DATA PROPROCESSING

INTRODUCTION:

In this initial phase, we will focus on the crucial steps of collecting and preprocessing historical stock market data. These steps lay the foundation for developing a predictive model in later stages.

DATA COLLECTION:

The first step in building a stock price prediction model is to gather historical stock market data. This data is essential for training and testing our model. For the purpose of this project, we will assume we have a CSV file named 'stock_data.csv' with columns like 'Date', 'Open', 'High', 'Low', 'Close', and 'Volume'.

CODE:

IMPORTING THE DATASET:

```
import pandas as pd  
df = pd.read_csv('MSFT.csv')  
df
```

DATA PREPROCESSING:

The data, the next step is data preprocessing. This stage involves several essential tasks:

CODE:

EXPLORE THE DATA:

```
import pandas as pd  
df = pd.read_csv('MSFT.csv')  
print(df.head(5))
```

```
import pandas as pd
df = pd.read_csv('MSFT.csv')
print(df.tail(5))
```

```
import pandas as pd
df = pd.read_csv('MSFT.csv')
print(df.info())
```

```
import pandas as pd
df = pd.read_csv('MSFT.csv')
print(df.describe())
```

HANDLE THE MISSING DATA:

```
import pandas as pd
df = pd.read_csv('MSFT.csv')
df.dropna(inplace=True)
```

```
import pandas as pd
df = pd.read_csv('MSFT.csv')
df.reset_index(drop=True, inplace=True)
```

VISUALIZE THE DATA:

```
import matplotlib.pyplot as plt
plt.figure(figsize=(16,8))
plt.title('Close Price History', fontsize=18)
plt.plot(df['Close'])
plt.xlabel('Date', fontsize=18)
plt.ylabel('Close Price USD ($)', fontsize=18)
plt.show()
```

SPLITTING THE DATA:

```
import math
import numpy as np
close_data = df.filter(['Close'])
dataset = close_data.values
training_data_len = math.ceil(len(dataset) * .8)
training_data_len
```

FEATURE SCALING:

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(dataset)
scaled_data
```

SPLITTING THE DATA:

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

```
x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
print(x_train.shape)

train_data = scaled_data[0:training_data_len, :]

x_train = []
y_train = []

for i in range(60, len(train_data)):
    x_train.append(train_data[i-60:i, 0])
    y_train.append(train_data[i, 0])

x_train, y_train = np.array(x_train), np.array(y_train)

print(x_train)
print(y_train)
```

CONCLUSION:

In this phase of building a stock price prediction model, we have focused on data collection and preprocessing. These early stages are vital for laying the groundwork for a successful predictive model.