copy-of-all-models-nikkei

November 10, 2024

[2]:
```
import pandas as pd
```

[47]:
```
import numpy as np
from sklearn.preprocessing import MinMaxScaler
import sklearn.metrics
```

[83]:
```
NSEI=pd.read_csv('/content/NIKKEI.csv')
```

[84]:
```
NSEI=NSEI.iloc[:,0:11]
```

[85]:
```
NSEI.shape
```

[85]:
```
(2616, 11)
```

[86]:
```
NSEI = NSEI.dropna(subset=['MA_50'])
```

[87]:
```
NSEI.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2517 entries, 49 to 2615
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Date          2517 non-null   object
 1   Open          2517 non-null   float64
 2   High          2517 non-null   float64
 3   Low           2517 non-null   float64
 4   Close         2517 non-null   float64
 5   Adj Close     2517 non-null   float64
 6   Volume        2517 non-null   float64
 7   MA_50         2517 non-null   float64
 8   Daily_Return  2517 non-null   float64
 9   Volatility    2517 non-null   float64
 10  Change        2517 non-null   float64
dtypes: float64(10), object(1)
memory usage: 236.0+ KB
```

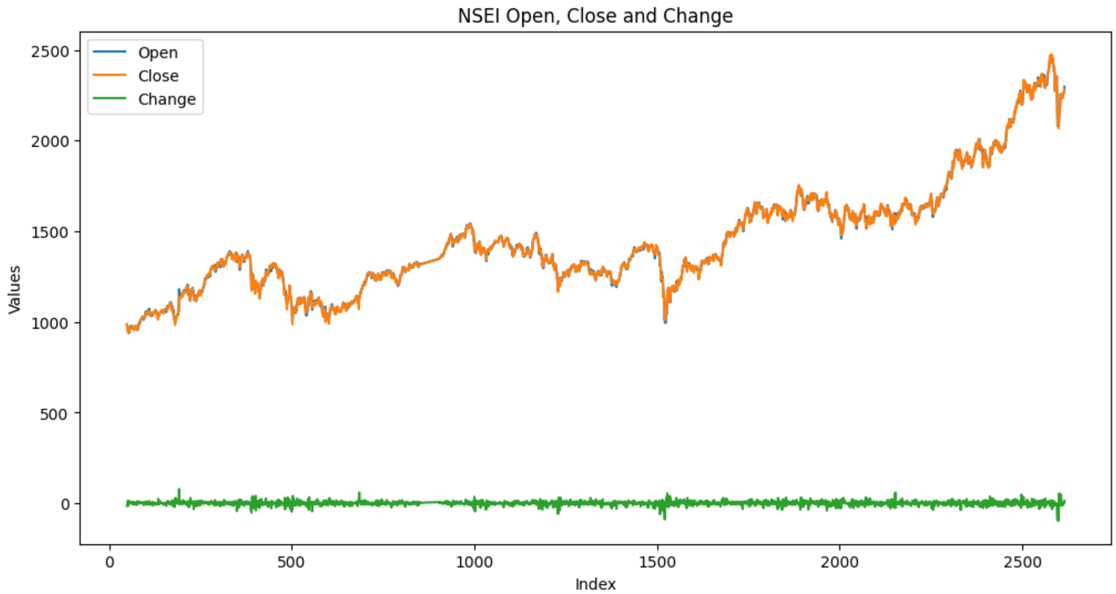
[88]:
```
NSEI['Date'] = pd.to_datetime(NSEI['Date']).dt.date
```

[89]:
```
NSEI.head()
```

[89]:
```
          Date   Open   High    Low  Close   Adj Close    Volume   MA_50  \
49  2014-04-07  985.0  993.0  984.0  988.0  793.799744    8983.0  981.52
50  2014-04-08  980.0  980.0  969.0  973.0  781.748352   26316.0  980.98
51  2014-04-09  966.0  966.0  946.0  954.0  766.482910  103949.0  980.06
52  2014-04-10  969.0  969.0  949.0  955.0  767.286255   37337.0  979.02
53  2014-04-11  942.0  948.0  931.0  941.0  756.038208   54703.0  977.36

    Daily_Return  Volatility  Change
49     -0.011011    0.011796   -14.0
50     -0.015182    0.012084    -8.0
51     -0.019527    0.012643    -7.0
52      0.001048    0.012495    15.0
53     -0.014660    0.012100   -13.0
```

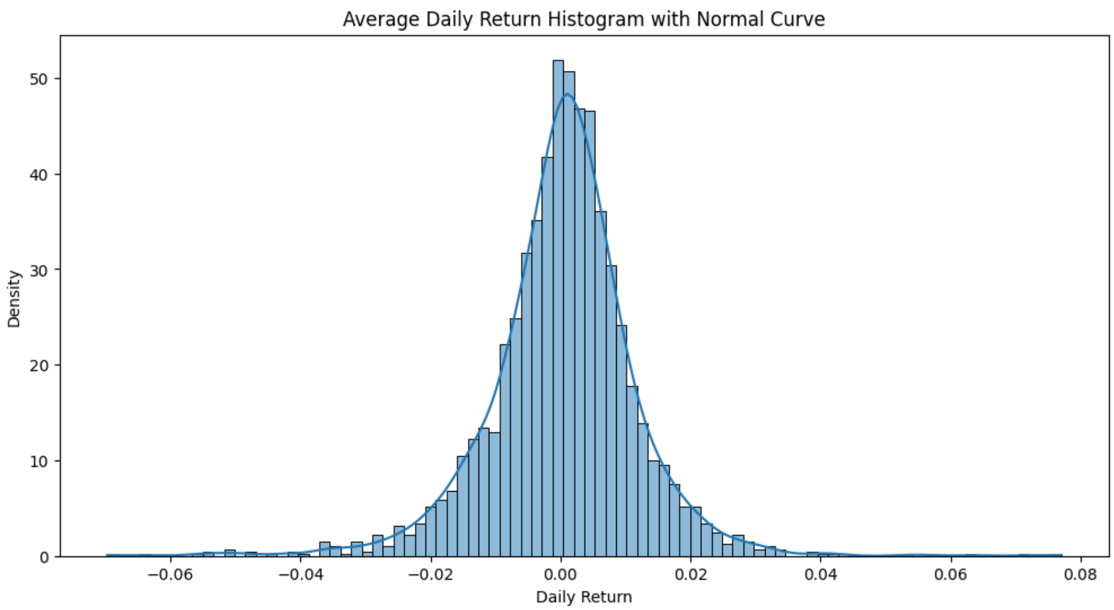
[90]:
```
gdata = NSEI[['Open', 'Close', 'Change']]
```

[91]:
```
import matplotlib.pyplot as plt

gdata.plot(figsize=(12, 6))
plt.title('NSEI Open, Close and Change')
plt.xlabel('Index')
plt.ylabel('Values')
plt.legend()
plt.show()
```



[92]:

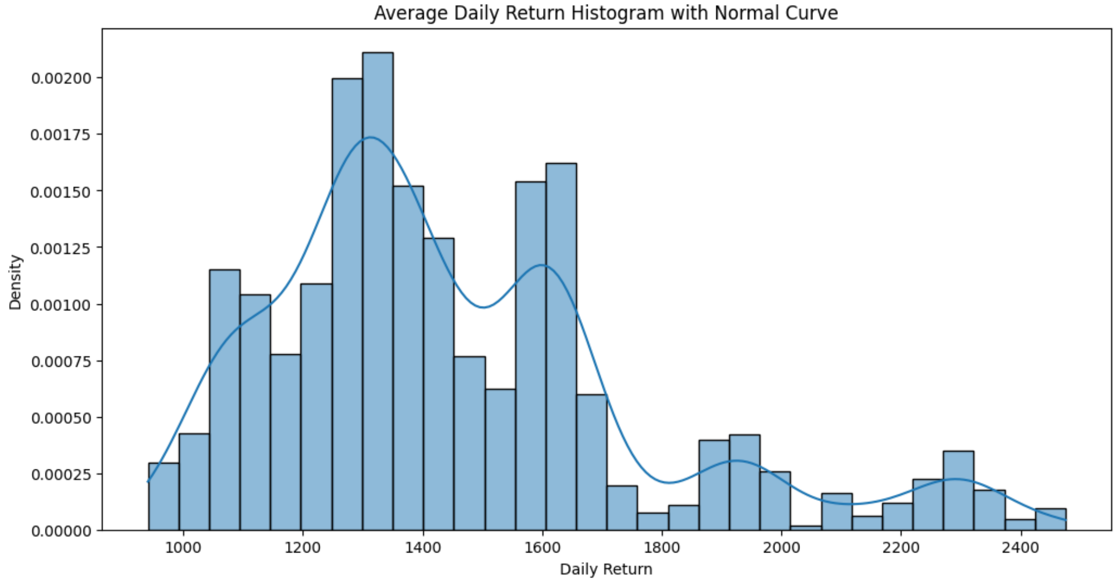
```
import numpy as np
import seaborn as sns
from scipy.stats import norm

plt.figure(figsize=(12, 6))
sns.histplot(NSEI['Daily_Return'].dropna(), kde=True, stat='density')

mu, std = norm.fit(NSEI['Daily_Return'].dropna())

xmin, xmax = plt.xlim()
x = np.linspace(xmin, xmax, 100)
plt.title('Average Daily Return Histogram with Normal Curve')
plt.xlabel('Daily Return')
plt.ylabel('Density')
plt.show()
```



[93]:

```
import numpy as np
import seaborn as sns
from scipy.stats import norm

plt.figure(figsize=(12, 6))
sns.histplot(NSEI['Close'].dropna(), kde=True, stat='density')

mu, std = norm.fit(NSEI['Close'].dropna())

xmin, xmax = plt.xlim()
x = np.linspace(xmin, xmax, 100)
```

```
plt.title('Average Daily Return Histogram with Normal Curve')
plt.xlabel('Daily Return')
plt.ylabel('Density')
plt.show()
```



[94]:
```
# prompt: Check outliers in NSEI. give count

def count_outliers_iqr(data):
  """Counts the number of outliers in a DataFrame using the IQR method.

  Args:
    data: A pandas DataFrame.

  Returns:
    A dictionary where keys are column names and values are the number of␣
↪outliers in each column.
  """
  outlier_counts = {}
  for column in data.columns:
    if pd.api.types.is_numeric_dtype(data[column]):
      Q1 = data[column].quantile(0.25)
      Q3 = data[column].quantile(0.75)
      IQR = Q3 - Q1
      lower_bound = Q1 - 1.5 * IQR
      upper_bound = Q3 + 1.5 * IQR
      outliers = data[(data[column] < lower_bound) | (data[column] >␣
↪upper_bound)]
```

```
      outlier_counts[column] = len(outliers)
  return outlier_counts


outlier_counts = count_outliers_iqr(NSEI)
print("Number of Outliers for Each Attribute:")
for column, count in outlier_counts.items():
  print(f"{column}: {count}")
```

```
Number of Outliers for Each Attribute:
Open: 133
High: 135
Low: 132
Close: 134
Adj Close: 132
Volume: 244
MA_50: 116
Daily_Return: 125
Volatility: 206
Change: 154
```

[95]:
```
def calculate_outlier_percentage(data):
  outlier_percentages = {}
  for column in data.columns:
    if pd.api.types.is_numeric_dtype(data[column]):
      Q1 = data[column].quantile(0.25)
      Q3 = data[column].quantile(0.75)
      IQR = Q3 - Q1
      lower_bound = Q1 - 1.5 * IQR
      upper_bound = Q3 + 1.5 * IQR
      outliers = data[(data[column] < lower_bound) | (data[column] >␣
 ↪upper_bound)]
      outlier_percentage = (len(outliers) / len(data)) * 100 if len(data) > 0␣
 ↪else 0
      outlier_percentages[column] = outlier_percentage

  return outlier_percentages
outlier_percentages = calculate_outlier_percentage(NSEI)
print("Outlier Percentages for Each Attribute:")
for column, percentage in outlier_percentages.items():
  print(f"{column}: {percentage:.2f}%")
```

```
Outlier Percentages for Each Attribute:
Open: 5.28%
High: 5.36%
Low: 5.24%
Close: 5.32%
```

```
Adj Close: 5.24%
Volume: 9.69%
MA_50: 4.61%
Daily_Return: 4.97%
Volatility: 8.18%
Change: 6.12%
```

[96]:
```python
def remove_outliers_iqr(df):
    # Calculate Q1 (25th percentile) and Q3 (75th percentile) for each column
    Q1 = df.quantile(0.25)
    Q3 = df.quantile(0.75)

    # Calculate the Interquartile Range (IQR)
    IQR = Q3 - Q1

    # Define the lower and upper bounds
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR

    # Remove outliers
    df_no_outliers_iqr = df[~((df < lower_bound) | (df > upper_bound)).
  ↪any(axis=1)]
    return df_no_outliers_iqr
NSEI = remove_outliers_iqr(NSEI)
```

ARIMA

[19]:
```python
#arima
NSEI['Close'] = np.log(NSEI['Close'])
```

[20]:
```python
NSEI.head()
```

[20]:
```
          Date   Open   High    Low     Close   Adj Close   Volume   MA_50  \
49  2014-04-07  985.0  993.0  984.0  6.895683  793.799744   8983.0  981.52
50  2014-04-08  980.0  980.0  969.0  6.880384  781.748352  26316.0  980.98
52  2014-04-10  969.0  969.0  949.0  6.861711  767.286255  37337.0  979.02
54  2014-04-14  938.0  948.0  937.0  6.846943  756.038208  13880.0  976.26
55  2014-04-15  953.0  954.0  940.0  6.846943  756.038208  12745.0  975.24

    Daily_Return  Volatility  Change
49     -0.011011    0.011796   -14.0
50     -0.015182    0.012084    -8.0
52      0.001048    0.012495    15.0
54      0.000000    0.012109    -3.0
55      0.000000    0.010030    12.0
```

[21]:
```python
!pip install pmdarima
```

```
Collecting pmdarima
  Downloading pmdarima-2.0.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86
_64.manylinux_2_28_x86_64.whl.metadata (7.8 kB)
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.10/dist-
packages (from pmdarima) (1.4.2)
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (3.0.11)
Requirement already satisfied: numpy>=1.21.2 in /usr/local/lib/python3.10/dist-
packages (from pmdarima) (1.26.4)
Requirement already satisfied: pandas>=0.19 in /usr/local/lib/python3.10/dist-
packages (from pmdarima) (2.2.2)
Requirement already satisfied: scikit-learn>=0.22 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (1.5.2)
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-
packages (from pmdarima) (1.13.1)
Requirement already satisfied: statsmodels>=0.13.2 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (0.14.4)
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-
packages (from pmdarima) (2.2.3)
Requirement already satisfied: setuptools!=50.0.0,>=38.6.0 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (75.1.0)
Requirement already satisfied: packaging>=17.1 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (24.1)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas>=0.19->pmdarima) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas>=0.19->pmdarima) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-
packages (from pandas>=0.19->pmdarima) (2024.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.22->pmdarima)
(3.5.0)
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.10/dist-
packages (from statsmodels>=0.13.2->pmdarima) (0.5.6)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages
(from patsy>=0.5.6->statsmodels>=0.13.2->pmdarima) (1.16.0)
Downloading pmdarima-2.0.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_6
4.manylinux_2_28_x86_64.whl (2.1 MB)
                          2.1/2.1 MB
20.7 MB/s eta 0:00:00
Installing collected packages: pmdarima
Successfully installed pmdarima-2.0.4
```

[22]:

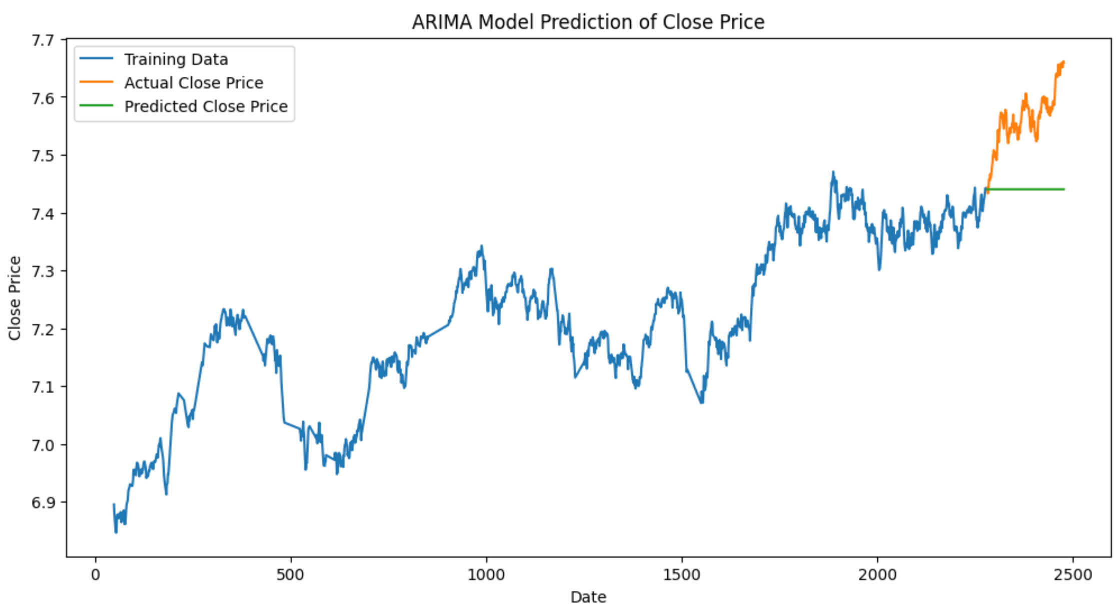
```
# prompt: write code for arima using autoarima to predict closing price in␣
↪given dataset. use training as 90% data

from pmdarima import auto_arima
```

```python
from sklearn.metrics import mean_squared_error

# Assuming 'NSEI' DataFrame is already loaded and prepared

# Split data into training and testing sets (90% train, 10% test)
train_data = NSEI['Close'][:-int(len(NSEI) * 0.1)]
test_data = NSEI['Close'][-int(len(NSEI) * 0.1):]

# Fit auto_arima model to the training data
model = auto_arima(train_data, start_p = 1, start_q = 1,
                          max_p = 100, max_q = 100,
                          start_P = 0,alpha=0.05,
                          trace = True,information_criterion='aic',
                          error_action ='ignore',   # we don't want to know if␣
 ↪an order does not work
                          suppress_warnings = True,  # we don't want␣
 ↪convergence warnings
                          stepwise = True)

# Make predictions on the test data
predictions = model.predict(n_periods=len(test_data))

# Evaluate the model
rmse = np.sqrt(mean_squared_error(test_data, predictions))
print(f'RMSE: {rmse}')

# Plot the results
plt.figure(figsize=(12, 6))
plt.plot(train_data, label='Training Data')
plt.plot(test_data.index, test_data, label='Actual Close Price')
plt.plot(test_data.index, predictions, label='Predicted Close Price')
plt.title('ARIMA Model Prediction of Close Price')
plt.xlabel('Date')
plt.ylabel('Close Price')
plt.legend()
plt.show()
```

```
Performing stepwise search to minimize aic
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=-10546.247, Time=2.34 sec
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=-10543.501, Time=0.59 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=-10548.144, Time=1.01 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=-10547.987, Time=2.32 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=-10543.963, Time=0.28 sec
 ARIMA(2,1,0)(0,0,0)[0] intercept   : AIC=-10546.289, Time=2.92 sec
 ARIMA(2,1,1)(0,0,0)[0] intercept   : AIC=-10544.243, Time=2.89 sec
 ARIMA(1,1,0)(0,0,0)[0]             : AIC=-10548.794, Time=0.50 sec
 ARIMA(2,1,0)(0,0,0)[0]             : AIC=-10546.967, Time=0.99 sec
```

```
 ARIMA(1,1,1)(0,0,0)[0]             : AIC=-10546.924, Time=1.83 sec
 ARIMA(0,1,1)(0,0,0)[0]             : AIC=-10548.622, Time=2.46 sec
 ARIMA(2,1,1)(0,0,0)[0]             : AIC=-10544.968, Time=2.46 sec

Best model:  ARIMA(1,1,0)(0,0,0)[0]
Total fit time: 20.675 seconds
RMSE: 0.1308410702835584
```

```
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/base/tsa_model.py:837:
ValueWarning: No supported index is available. Prediction results will be given
with an integer index beginning at `start`.
  return get_prediction_index(
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/base/tsa_model.py:837:
FutureWarning: No supported index is available. In the next version, calling
this method in a model without a supported index will result in an exception.
  return get_prediction_index(
```



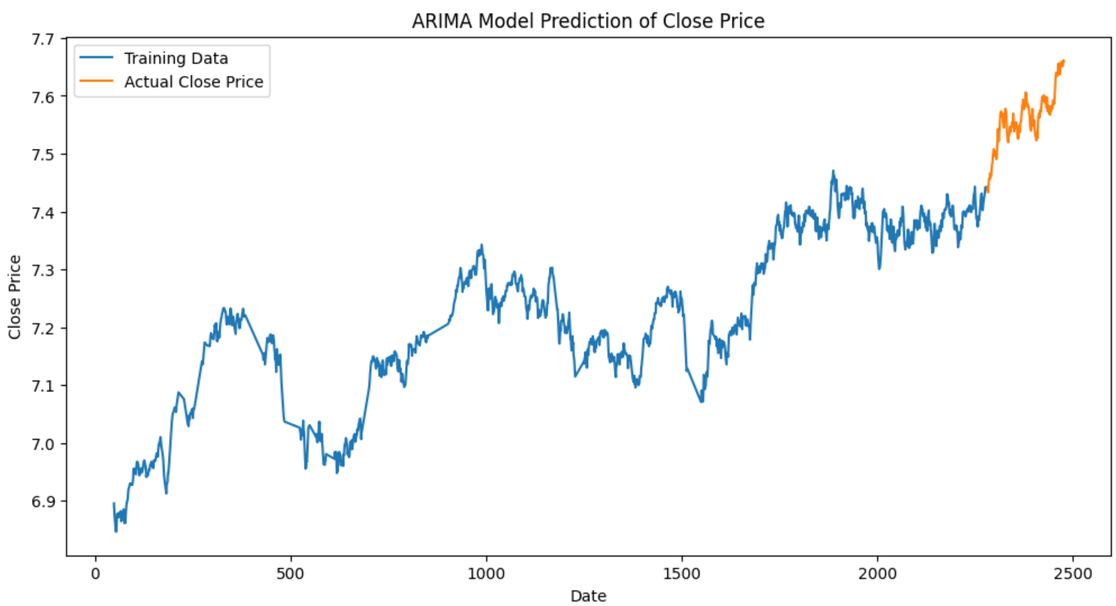
[23]:
```
# prompt: find out confusion matrix,rmse,mse and other evaluation matrics for
 ↪the above fir

from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score,
 ↪recall_score, f1_score

# Assuming 'test_data' and 'predictions' are already defined from the previous
 ↪code
```

```python
# Calculate RMSE (Root Mean Squared Error)
rmse = np.sqrt(mean_squared_error(test_data, predictions))

# Calculate MSE (Mean Squared Error)
mse = mean_squared_error(test_data, predictions)

# Calculate MAE (Mean Absolute Error)
mae = mean_absolute_error(test_data, predictions)

# Calculate R-squared
r2 = r2_score(test_data, predictions)

print(f'RMSE: {rmse}')
print(f'MSE: {mse}')
print(f'MAE: {mae}')
print(f'R-squared: {r2}')

# You can also calculate other metrics like MAPE (Mean Absolute Percentage
 ↪Error)
# if needed, but it might require some custom implementation.

# For Classification metrics, you'd need to convert your predictions into
 ↪discrete classes
# (e.g., based on a threshold) and then calculate things like confusion matrix,
# accuracy, precision, recall, F1-score.


# Example of converting predictions to binary classes (assuming a threshold of
 ↪0.5):
# predicted_classes = (predictions > 0.5).astype(int)
# actual_classes = (test_data > 0.5).astype(int)

# Calculate confusion matrix
# cm = confusion_matrix(actual_classes, predicted_classes)
# print("Confusion Matrix:\n", cm)

# Calculate accuracy
# accuracy = accuracy_score(actual_classes, predicted_classes)
# print("Accuracy:", accuracy)

# Calculate precision
# precision = precision_score(actual_classes, predicted_classes)
# print("Precision:", precision)

# Calculate recall
# recall = recall_score(actual_classes, predicted_classes)
# print("Recall:", recall)
```

```
# Calculate F1-score
# f1 = f1_score(actual_classes, predicted_classes)
# print("F1-score:", f1)
```

```
RMSE: 0.1308410702835584
MSE: 0.017119385672947067
MAE: 0.12185498943436865
R-squared: -6.437206148298887
```

[24]:
```
# prompt: # prompt: generate the  training testing graph

# Assuming 'train_data', 'test_data', and 'predictions' are already defined

plt.figure(figsize=(12, 6))
plt.plot(train_data, label='Training Data')
plt.plot(test_data.index, test_data, label='Actual Close Price')
plt.title('ARIMA Model Prediction of Close Price')
plt.xlabel('Date')
plt.ylabel('Close Price')
plt.legend()
plt.show()
```



[25]:
```
# prompt: check for overfitting in above model

# Calculate the training and testing RMSE
train_predictions = model.predict_in_sample()
```

```
train_rmse = np.sqrt(mean_squared_error(train_data, train_predictions))
test_rmse = np.sqrt(mean_squared_error(test_data, predictions))

print(f"Training RMSE: {train_rmse}")
print(f"Testing RMSE: {test_rmse}")

# Check for overfitting by comparing training and testing RMSE
if test_rmse > train_rmse and (test_rmse - train_rmse) > some_threshold:  # You␣
↪can define a threshold for significance
  print("Warning: The model might be overfitting.")
  print("The testing RMSE is considerably higher than the training RMSE,␣
↪indicating the model is performing poorly on unseen data.")
else:
  print("The model doesn't appear to be overfitting significantly.")
```

```
Training RMSE: 0.16802775924390476
Testing RMSE: 0.1308410702835584
The model doesn't appear to be overfitting significantly.
```

[26]:

```
# Calculate the training and testing RMSE
train_predictions = model.predict_in_sample()
train_rmse = np.sqrt(mean_squared_error(train_data, train_predictions))
test_rmse = np.sqrt(mean_squared_error(test_data, predictions))

print(f"Training RMSE: {train_rmse}")
print(f"Testing RMSE: {test_rmse}")

# Check for underfitting by comparing training and testing RMSE and the␣
↪baseline RMSE
# Create an array of baseline predictions with the same length as test_data
baseline_predictions = np.repeat(np.mean(train_data), len(test_data))  # Repeat␣
↪the mean for each test data point
baseline_rmse = np.sqrt(mean_squared_error(test_data, baseline_predictions))  #␣
↪Calculate RMSE using baseline predictions
print(f"Baseline RMSE: {baseline_rmse}")

if train_rmse > baseline_rmse and test_rmse > baseline_rmse:
    print("Warning: The model might be underfitting.")
    print("Both training and testing RMSE are higher than the baseline RMSE,␣
↪indicating the model is not learning effectively.")
elif train_rmse < baseline_rmse and test_rmse > baseline_rmse:
    print("The model is performing better than the baseline on the training␣
↪data but not on the testing data.")
    print("This might indicate that it's not generalizing well or that the␣
↪training data is not representative enough.")
else:
```

```
    print("The model doesn't appear to be underfitting significantly.")

# You can also consider the R-squared value as another indicator for
↪underfitting.
# A low R-squared value (e.g., close to 0) suggests that the model is not
↪explaining much of the variance in the data.
```

```
Training RMSE: 0.16802775924390476
Testing RMSE: 0.1308410702835584
Baseline RMSE: 0.34012156680825734
The model doesn't appear to be underfitting significantly.
```

GRU

[42]:
```
!pip install tensorflow
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import GRU, Dense
```

```
Requirement already satisfied: tensorflow in /usr/local/lib/python3.10/dist-
packages (2.17.0)
Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.10/dist-
packages (from tensorflow) (1.4.0)
Requirement already satisfied: astunparse>=1.6.0 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (1.6.3)
Requirement already satisfied: flatbuffers>=24.3.25 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (24.3.25)
Requirement already satisfied: gast!=0.5.0,!=0.5.1,!=0.5.2,>=0.2.1 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (0.6.0)
Requirement already satisfied: google-pasta>=0.1.1 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (0.2.0)
Requirement already satisfied: h5py>=3.10.0 in /usr/local/lib/python3.10/dist-
packages (from tensorflow) (3.12.1)
Requirement already satisfied: libclang>=13.0.0 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (18.1.1)
Requirement already satisfied: ml-dtypes<0.5.0,>=0.3.1 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (0.4.1)
Requirement already satisfied: opt-einsum>=2.3.2 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (3.4.0)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-
packages (from tensorflow) (24.1)
Requirement already satisfied:
protobuf!=4.21.0,!=4.21.1,!=4.21.2,!=4.21.3,!=4.21.4,!=4.21.5,<5.0.0dev,>=3.20.3
in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.20.3)
Requirement already satisfied: requests<3,>=2.21.0 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (2.32.3)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-
packages (from tensorflow) (75.1.0)
Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.10/dist-
```

```
packages (from tensorflow) (1.16.0)
Requirement already satisfied: termcolor>=1.1.0 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (2.5.0)
Requirement already satisfied: typing-extensions>=3.6.6 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (4.12.2)
Requirement already satisfied: wrapt>=1.11.0 in /usr/local/lib/python3.10/dist-
packages (from tensorflow) (1.16.0)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (1.64.1)
Requirement already satisfied: tensorboard<2.18,>=2.17 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (2.17.0)
Requirement already satisfied: keras>=3.2.0 in /usr/local/lib/python3.10/dist-
packages (from tensorflow) (3.4.1)
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (0.37.1)
Requirement already satisfied: numpy<2.0.0,>=1.23.5 in
/usr/local/lib/python3.10/dist-packages (from tensorflow) (1.26.4)
Requirement already satisfied: wheel<1.0,>=0.23.0 in
/usr/local/lib/python3.10/dist-packages (from astunparse>=1.6.0->tensorflow)
(0.44.0)
Requirement already satisfied: rich in /usr/local/lib/python3.10/dist-packages
(from keras>=3.2.0->tensorflow) (13.9.3)
Requirement already satisfied: namex in /usr/local/lib/python3.10/dist-packages
(from keras>=3.2.0->tensorflow) (0.0.8)
Requirement already satisfied: optree in /usr/local/lib/python3.10/dist-packages
(from keras>=3.2.0->tensorflow) (0.13.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorflow)
(3.4.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests<3,>=2.21.0->tensorflow) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorflow)
(2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorflow)
(2024.8.30)
Requirement already satisfied: markdown>=2.6.8 in
/usr/local/lib/python3.10/dist-packages (from
tensorboard<2.18,>=2.17->tensorflow) (3.7)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in
/usr/local/lib/python3.10/dist-packages (from
tensorboard<2.18,>=2.17->tensorflow) (0.7.2)
Requirement already satisfied: werkzeug>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from
tensorboard<2.18,>=2.17->tensorflow) (3.0.6)
Requirement already satisfied: MarkupSafe>=2.1.1 in
/usr/local/lib/python3.10/dist-packages (from
```

```
werkzeug>=1.0.1->tensorboard<2.18,>=2.17->tensorflow) (3.0.2)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.10/dist-packages (from rich->keras>=3.2.0->tensorflow)
(3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.10/dist-packages (from rich->keras>=3.2.0->tensorflow)
(2.18.0)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-
packages (from markdown-it-py>=2.2.0->rich->keras>=3.2.0->tensorflow) (0.1.2)
```

[43]:
```
# prompt: # prompt: write a code to apply GRU on NSEI dataset to predict close␣
↪using all attributes. use min max scalar for pre-processing on all numeric␣
↪attributes. use 90% training data and 10% testing data.

# Assuming 'NSEI' DataFrame is already loaded and prepared
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt
# Extract relevant features for prediction (all attributes except 'Date')
data = NSEI.drop('Date', axis=1)

# Normalize the data using MinMaxScaler for numeric attributes
scaler = MinMaxScaler()
numeric_cols = data.select_dtypes(include=np.number).columns
data[numeric_cols] = scaler.fit_transform(data[numeric_cols])

# Split the data into training and testing sets (90% train, 10% test)
train_size = int(len(data) * 0.90)
test_size = len(data) - train_size
train_data, test_data = data[0:train_size], data[train_size:len(data)]

# Separate the 'Close' column as the target variable for both train and test␣
↪sets
trainY = train_data['Close'].values
trainX = train_data.drop('Close', axis=1).values
testY = test_data['Close'].values
testX = test_data.drop('Close', axis=1).values



# Reshape input to be [samples, time steps, features]
trainX = np.reshape(trainX, (trainX.shape[0], 1, trainX.shape[1]))
testX = np.reshape(testX, (testX.shape[0], 1, testX.shape[1]))

# Create and fit the GRU network
model = Sequential()
model.add(GRU(units=50, return_sequences=True, input_shape=(trainX.shape[1],␣
↪trainX.shape[2])))
model.add(GRU(units=50))
```

```python
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(trainX, trainY, epochs=100, batch_size=32, verbose=2)

# Make predictions
trainPredict = model.predict(trainX)
testPredict = model.predict(testX)


# Invert predictions back to original scale for 'Close' column
trainPredict = scaler.inverse_transform(np.concatenate((np.zeros((trainPredict.
    ↪shape[0], data.shape[1] - 1)), trainPredict), axis=1))[:, -1]
trainY = scaler.inverse_transform(np.concatenate((np.zeros((trainY.shape[0],␣
    ↪data.shape[1] - 1)), trainY.reshape(-1, 1)), axis=1))[:, -1]

testPredict = scaler.inverse_transform(np.concatenate((np.zeros((testPredict.
    ↪shape[0], data.shape[1] - 1)), testPredict), axis=1))[:, -1]
testY = scaler.inverse_transform(np.concatenate((np.zeros((testY.shape[0], data.
    ↪shape[1] - 1)), testY.reshape(-1, 1)), axis=1))[:, -1]


# Calculate root mean squared error
trainScore = np.sqrt(mean_squared_error(trainY, trainPredict))
print('Train Score: %.2f RMSE' % (trainScore))
testScore = np.sqrt(mean_squared_error(testY, testPredict))
print('Test Score: %.2f RMSE' % (testScore))

# Plot the results
plt.figure(figsize=(12, 6))
plt.plot(NSEI.index[:train_size], trainY, label='Training Actual Close')
plt.plot(NSEI.index[train_size:], testY, label='Testing Actual Close')
plt.plot(NSEI.index[train_size:], testPredict, label='Testing Predicted Close')
plt.title('GRU Model Prediction of Close Price')
plt.xlabel('Date')
plt.ylabel('Close Price')
plt.legend()
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/keras/src/layers/rnn/rnn.py:204:
UserWarning: Do not pass an `input_shape`/`input_dim` argument to a layer. When
using Sequential models, prefer using an `Input(shape)` object as the first
layer in the model instead.
  super().__init__(**kwargs)
```

```
Epoch 1/100
53/53 - 12s - 222ms/step - loss: 0.0203
Epoch 2/100
53/53 - 1s - 20ms/step - loss: 4.8516e-04
```

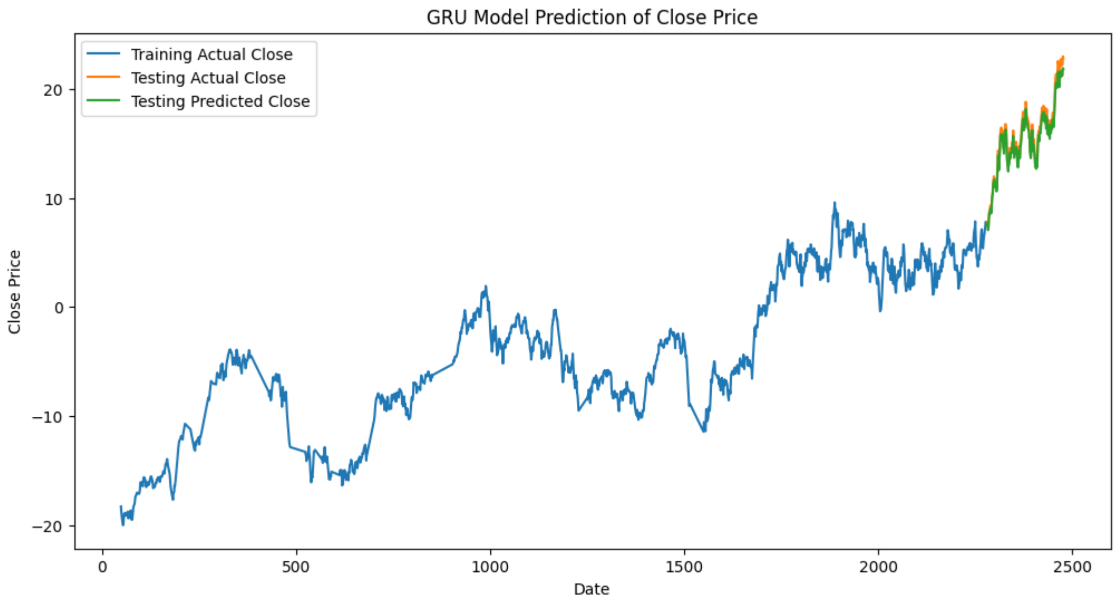
```
Epoch 3/100
53/53 - 1s - 12ms/step - loss: 1.9355e-04
Epoch 4/100
53/53 - 1s - 11ms/step - loss: 1.6036e-04
Epoch 5/100
53/53 - 0s - 7ms/step - loss: 1.3600e-04
Epoch 6/100
53/53 - 1s - 12ms/step - loss: 1.2763e-04
Epoch 7/100
53/53 - 1s - 13ms/step - loss: 1.0977e-04
Epoch 8/100
53/53 - 1s - 11ms/step - loss: 1.0210e-04
Epoch 9/100
53/53 - 1s - 14ms/step - loss: 9.1575e-05
Epoch 10/100
53/53 - 1s - 24ms/step - loss: 8.2117e-05
Epoch 11/100
53/53 - 0s - 9ms/step - loss: 7.1310e-05
Epoch 12/100
53/53 - 1s - 14ms/step - loss: 6.7115e-05
Epoch 13/100
53/53 - 1s - 10ms/step - loss: 5.5648e-05
Epoch 14/100
53/53 - 1s - 12ms/step - loss: 5.0437e-05
Epoch 15/100
53/53 - 1s - 11ms/step - loss: 4.4342e-05
Epoch 16/100
53/53 - 1s - 11ms/step - loss: 3.8844e-05
Epoch 17/100
53/53 - 0s - 9ms/step - loss: 3.3817e-05
Epoch 18/100
53/53 - 0s - 6ms/step - loss: 3.0076e-05
Epoch 19/100
53/53 - 1s - 13ms/step - loss: 2.6972e-05
Epoch 20/100
53/53 - 0s - 8ms/step - loss: 2.2344e-05
Epoch 21/100
53/53 - 0s - 7ms/step - loss: 2.0763e-05
Epoch 22/100
53/53 - 1s - 10ms/step - loss: 1.7449e-05
Epoch 23/100
53/53 - 0s - 8ms/step - loss: 1.5079e-05
Epoch 24/100
53/53 - 1s - 11ms/step - loss: 1.3315e-05
Epoch 25/100
53/53 - 0s - 8ms/step - loss: 1.1323e-05
Epoch 26/100
53/53 - 0s - 6ms/step - loss: 1.0085e-05
```

```
Epoch 27/100
53/53 - 0s - 6ms/step - loss: 9.3771e-06
Epoch 28/100
53/53 - 0s - 5ms/step - loss: 8.2744e-06
Epoch 29/100
53/53 - 0s - 6ms/step - loss: 7.3108e-06
Epoch 30/100
53/53 - 0s - 6ms/step - loss: 6.3091e-06
Epoch 31/100
53/53 - 0s - 6ms/step - loss: 6.0222e-06
Epoch 32/100
53/53 - 0s - 5ms/step - loss: 6.1868e-06
Epoch 33/100
53/53 - 0s - 4ms/step - loss: 5.1900e-06
Epoch 34/100
53/53 - 0s - 6ms/step - loss: 5.1640e-06
Epoch 35/100
53/53 - 0s - 4ms/step - loss: 4.3874e-06
Epoch 36/100
53/53 - 0s - 6ms/step - loss: 4.3351e-06
Epoch 37/100
53/53 - 0s - 4ms/step - loss: 4.0660e-06
Epoch 38/100
53/53 - 0s - 5ms/step - loss: 4.3563e-06
Epoch 39/100
53/53 - 0s - 3ms/step - loss: 3.5206e-06
Epoch 40/100
53/53 - 0s - 5ms/step - loss: 3.6753e-06
Epoch 41/100
53/53 - 0s - 3ms/step - loss: 4.3834e-06
Epoch 42/100
53/53 - 0s - 6ms/step - loss: 3.5907e-06
Epoch 43/100
53/53 - 0s - 6ms/step - loss: 3.7333e-06
Epoch 44/100
53/53 - 0s - 3ms/step - loss: 3.5678e-06
Epoch 45/100
53/53 - 0s - 7ms/step - loss: 3.2545e-06
Epoch 46/100
53/53 - 0s - 6ms/step - loss: 3.2345e-06
Epoch 47/100
53/53 - 0s - 6ms/step - loss: 3.4223e-06
Epoch 48/100
53/53 - 0s - 6ms/step - loss: 3.1655e-06
Epoch 49/100
53/53 - 0s - 6ms/step - loss: 3.8307e-06
Epoch 50/100
53/53 - 0s - 6ms/step - loss: 3.6234e-06
```

```
Epoch 51/100
53/53 - 0s - 6ms/step - loss: 3.4031e-06
Epoch 52/100
53/53 - 0s - 6ms/step - loss: 3.3801e-06
Epoch 53/100
53/53 - 0s - 6ms/step - loss: 3.7741e-06
Epoch 54/100
53/53 - 1s - 11ms/step - loss: 3.4719e-06
Epoch 55/100
53/53 - 0s - 4ms/step - loss: 3.6317e-06
Epoch 56/100
53/53 - 0s - 5ms/step - loss: 3.6724e-06
Epoch 57/100
53/53 - 0s - 4ms/step - loss: 3.0499e-06
Epoch 58/100
53/53 - 0s - 6ms/step - loss: 3.6502e-06
Epoch 59/100
53/53 - 0s - 4ms/step - loss: 3.8879e-06
Epoch 60/100
53/53 - 0s - 4ms/step - loss: 3.2574e-06
Epoch 61/100
53/53 - 0s - 6ms/step - loss: 4.8145e-06
Epoch 62/100
53/53 - 0s - 4ms/step - loss: 4.6973e-06
Epoch 63/100
53/53 - 0s - 4ms/step - loss: 3.5608e-06
Epoch 64/100
53/53 - 0s - 5ms/step - loss: 3.2072e-06
Epoch 65/100
53/53 - 0s - 4ms/step - loss: 3.3636e-06
Epoch 66/100
53/53 - 0s - 4ms/step - loss: 3.3231e-06
Epoch 67/100
53/53 - 0s - 4ms/step - loss: 3.4421e-06
Epoch 68/100
53/53 - 0s - 6ms/step - loss: 5.0873e-06
Epoch 69/100
53/53 - 0s - 6ms/step - loss: 3.5421e-06
Epoch 70/100
53/53 - 0s - 6ms/step - loss: 4.0502e-06
Epoch 71/100
53/53 - 0s - 3ms/step - loss: 3.5005e-06
Epoch 72/100
53/53 - 0s - 5ms/step - loss: 3.5915e-06
Epoch 73/100
53/53 - 0s - 3ms/step - loss: 5.5267e-06
Epoch 74/100
53/53 - 0s - 4ms/step - loss: 3.7296e-06
```

```
Epoch 75/100
53/53 - 0s - 6ms/step - loss: 3.3241e-06
Epoch 76/100
53/53 - 0s - 5ms/step - loss: 3.3693e-06
Epoch 77/100
53/53 - 0s - 6ms/step - loss: 3.3214e-06
Epoch 78/100
53/53 - 0s - 6ms/step - loss: 5.5218e-06
Epoch 79/100
53/53 - 0s - 3ms/step - loss: 4.1880e-06
Epoch 80/100
53/53 - 0s - 3ms/step - loss: 3.0481e-06
Epoch 81/100
53/53 - 0s - 6ms/step - loss: 4.5905e-06
Epoch 82/100
53/53 - 0s - 6ms/step - loss: 5.1271e-06
Epoch 83/100
53/53 - 0s - 3ms/step - loss: 3.0066e-06
Epoch 84/100
53/53 - 0s - 6ms/step - loss: 3.2212e-06
Epoch 85/100
53/53 - 0s - 3ms/step - loss: 7.2026e-06
Epoch 86/100
53/53 - 0s - 4ms/step - loss: 4.4602e-06
Epoch 87/100
53/53 - 0s - 3ms/step - loss: 3.2333e-06
Epoch 88/100
53/53 - 0s - 3ms/step - loss: 3.2817e-06
Epoch 89/100
53/53 - 0s - 6ms/step - loss: 4.4940e-06
Epoch 90/100
53/53 - 0s - 3ms/step - loss: 3.1376e-06
Epoch 91/100
53/53 - 0s - 6ms/step - loss: 3.7896e-06
Epoch 92/100
53/53 - 0s - 5ms/step - loss: 3.4829e-06
Epoch 93/100
53/53 - 0s - 6ms/step - loss: 8.4074e-06
Epoch 94/100
53/53 - 0s - 6ms/step - loss: 4.1447e-06
Epoch 95/100
53/53 - 0s - 6ms/step - loss: 5.5316e-06
Epoch 96/100
53/53 - 0s - 6ms/step - loss: 5.9487e-06
Epoch 97/100
53/53 - 0s - 6ms/step - loss: 4.7269e-06
Epoch 98/100
53/53 - 0s - 6ms/step - loss: 5.9852e-06
```

```
Epoch 99/100
53/53 - 0s - 5ms/step - loss: 2.8678e-06
Epoch 100/100
53/53 - 0s - 6ms/step - loss: 7.0512e-06
53/53              1s 13ms/step
6/6              0s 4ms/step
Train Score: 0.12 RMSE
Test Score: 0.56 RMSE
```



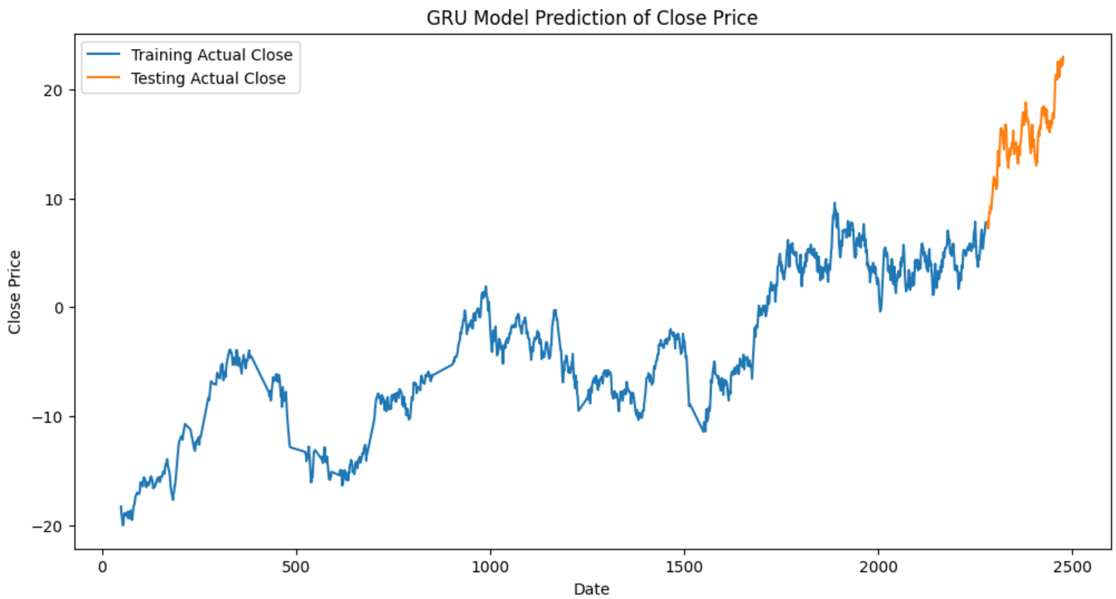
[44]:

```
# prompt: # prompt: find out confusion matrix,rmse,mse and other evaluation␣
↪matrics for the above model
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score #␣
↪Import mean_absolute_error,r2_score

# Assuming 'test_data' and 'predictions' are already defined from the previous␣
↪code

# Calculate RMSE (Root Mean Squared Error)
rmse = np.sqrt(mean_squared_error(testY,testPredict))

# Calculate MSE (Mean Squared Error)
mse = mean_squared_error(testY,testPredict)

# Calculate MAE (Mean Absolute Error)
mae = mean_absolute_error(testY,testPredict)

# Calculate R-squared
```

```
r2 = r2_score(testY,testPredict)

print(f'RMSE: {rmse}')
print(f'MSE: {mse}')
print(f'MAE: {mae}')
print(f'R-squared: {r2}')
```

```
RMSE: 0.5551042160522027
MSE: 0.3081406906789305
MAE: 0.5138690289733334
R-squared: 0.9730823738080759
```

[45]:
```
plt.figure(figsize=(12, 6))
plt.plot(NSEI.index[:train_size], trainY, label='Training Actual Close')
plt.plot(NSEI.index[train_size:], testY, label='Testing Actual Close')
plt.title('GRU Model Prediction of Close Price')
plt.xlabel('Date')
plt.ylabel('Close Price')
plt.legend()
plt.show()
```



[46]:
```
# Calculate the training and testing RMSE
trainScore = np.sqrt(mean_squared_error(trainY, trainPredict))
testScore = np.sqrt(mean_squared_error(testY, testPredict))

print(f"Training RMSE: {trainScore}")
print(f"Testing RMSE: {testScore}")
```

```python
# Define a threshold for significance
some_threshold = 5  # You can adjust this value based on your data and model

# Check for overfitting by comparing training and testing RMSE
if testScore > trainScore and (testScore - trainScore) > some_threshold:  # You␣
↪can define a threshold for significance
  print("Warning: The model might be overfitting.")
  print("The testing RMSE is considerably higher than the training RMSE,␣
↪indicating the model is performing poorly on unseen data.")
else:
  print("The model doesn't appear to be overfitting significantly.")


# Check for underfitting by comparing training and testing RMSE and the␣
↪baseline RMSE
# Create an array of baseline predictions with the same length as test_data
baseline_predictions = np.repeat(np.mean(trainY), len(testY))  # Repeat the␣
↪mean for each test data point
baseline_rmse = np.sqrt(mean_squared_error(testY, baseline_predictions))  #␣
↪Calculate RMSE using baseline predictions
print(f"Baseline RMSE: {baseline_rmse}")

if trainScore > baseline_rmse and testScore > baseline_rmse:
    print("Warning: The model might be underfitting.")
    print("Both training and testing RMSE are higher than the baseline RMSE,␣
↪indicating the model is not learning effectively.")
elif trainScore < baseline_rmse and testScore > baseline_rmse:
    print("The model is performing better than the baseline on the training␣
↪data but not on the testing data.")
    print("This might indicate that it's not generalizing well or that the␣
↪training data is not representative enough.")
else:
    print("The model doesn't appear to be underfitting significantly.")
```

```
Training RMSE: 0.1221832295241226
Testing RMSE: 0.5551042160522027
The model doesn't appear to be overfitting significantly.
Baseline RMSE: 19.819334804725994
The model doesn't appear to be underfitting significantly.
```

LSTM

[62]:
```python
# prompt: write a code to apply LSTM on NSEI dataset to predict close using all␣
↪attributes. use min max scalar for pre-processing on all numeric attributes.␣
↪use 90% training data and 10% testing data.
```

```python
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
from sklearn.metrics import mean_squared_error

# Assuming 'NSEI' DataFrame is already loaded and prepared
# Extract relevant features for prediction (all attributes except 'Date')
data = NSEI.drop('Date', axis=1)

# Normalize the data using MinMaxScaler for numeric attributes
scaler = MinMaxScaler()
numeric_cols = data.select_dtypes(include=np.number).columns
data[numeric_cols] = scaler.fit_transform(data[numeric_cols])

# Split the data into training and testing sets (90% train, 10% test)
train_size = int(len(data) * 0.90)
test_size = len(data) - train_size
train_data, test_data = data[0:train_size], data[train_size:len(data)]

# Separate the 'Close' column as the target variable for both train and test␣
↪sets
trainY = train_data['Close'].values
trainX = train_data.drop('Close', axis=1).values
testY = test_data['Close'].values
testX = test_data.drop('Close', axis=1).values

# Reshape input to be [samples, time steps, features]
trainX = np.reshape(trainX, (trainX.shape[0], 1, trainX.shape[1]))
testX = np.reshape(testX, (testX.shape[0], 1, testX.shape[1]))

# Create and fit the LSTM network
model = Sequential()
model.add(LSTM(units=50, return_sequences=True, input_shape=(trainX.shape[1],␣
↪trainX.shape[2])))
model.add(LSTM(units=50))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(trainX, trainY, epochs=100, batch_size=32, verbose=2)

# Make predictions
trainPredict = model.predict(trainX)
testPredict = model.predict(testX)

# Invert predictions back to original scale for 'Close' column
```

```
trainPredict = scaler.inverse_transform(np.concatenate((np.zeros((trainPredict.
↪shape[0], data.shape[1] - 1)), trainPredict), axis=1))[:, -1]
trainY = scaler.inverse_transform(np.concatenate((np.zeros((trainY.shape[0],␣
↪data.shape[1] - 1)), trainY.reshape(-1, 1)), axis=1))[:, -1]

testPredict = scaler.inverse_transform(np.concatenate((np.zeros((testPredict.
↪shape[0], data.shape[1] - 1)), testPredict), axis=1))[:, -1]
testY = scaler.inverse_transform(np.concatenate((np.zeros((testY.shape[0], data.
↪shape[1] - 1)), testY.reshape(-1, 1)), axis=1))[:, -1]

# Calculate root mean squared error
trainScore = np.sqrt(mean_squared_error(trainY, trainPredict))
print('Train Score: %.2f RMSE' % (trainScore))
testScore = np.sqrt(mean_squared_error(testY, testPredict))
print('Test Score: %.2f RMSE' % (testScore))

# Plot the results
import matplotlib.pyplot as plt
plt.figure(figsize=(12, 6))
plt.plot(NSEI.index[:train_size], trainY, label='Training Actual Close')
plt.plot(NSEI.index[train_size:], testY, label='Testing Actual Close')
plt.plot(NSEI.index[train_size:], testPredict, label='Testing Predicted Close')
plt.title('LSTM Model Prediction of Close Price')
plt.xlabel('Date')
plt.ylabel('Close Price')
plt.legend()
plt.show()
```

```
Epoch 1/100
```

```
/usr/local/lib/python3.10/dist-packages/keras/src/layers/rnn/rnn.py:204:
UserWarning: Do not pass an `input_shape`/`input_dim` argument to a layer. When
using Sequential models, prefer using an `Input(shape)` object as the first
layer in the model instead.
  super().__init__(**kwargs)
```

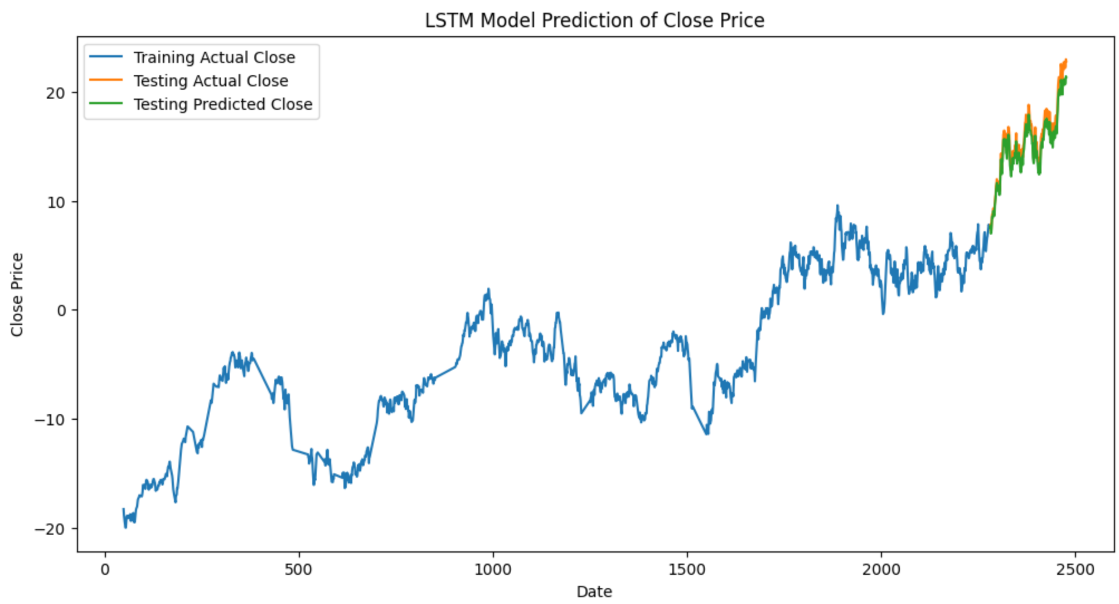
```
53/53 - 3s - 63ms/step - loss: 0.0562
Epoch 2/100
53/53 - 0s - 5ms/step - loss: 0.0041
Epoch 3/100
53/53 - 0s - 6ms/step - loss: 6.0439e-04
Epoch 4/100
53/53 - 0s - 5ms/step - loss: 3.1540e-04
Epoch 5/100
53/53 - 0s - 3ms/step - loss: 2.4236e-04
Epoch 6/100
53/53 - 0s - 3ms/step - loss: 2.0597e-04
Epoch 7/100
```

```
53/53 - 0s - 6ms/step - loss: 1.7722e-04
Epoch 8/100
53/53 - 0s - 4ms/step - loss: 1.5771e-04
Epoch 9/100
53/53 - 0s - 5ms/step - loss: 1.4049e-04
Epoch 10/100
53/53 - 0s - 6ms/step - loss: 1.2586e-04
Epoch 11/100
53/53 - 0s - 5ms/step - loss: 1.1419e-04
Epoch 12/100
53/53 - 0s - 6ms/step - loss: 9.9054e-05
Epoch 13/100
53/53 - 0s - 7ms/step - loss: 8.7846e-05
Epoch 14/100
53/53 - 0s - 7ms/step - loss: 8.0369e-05
Epoch 15/100
53/53 - 0s - 6ms/step - loss: 7.3838e-05
Epoch 16/100
53/53 - 0s - 5ms/step - loss: 6.7150e-05
Epoch 17/100
53/53 - 0s - 6ms/step - loss: 6.3028e-05
Epoch 18/100
53/53 - 1s - 12ms/step - loss: 5.7600e-05
Epoch 19/100
53/53 - 1s - 12ms/step - loss: 5.0960e-05
Epoch 20/100
53/53 - 0s - 4ms/step - loss: 4.7361e-05
Epoch 21/100
53/53 - 0s - 4ms/step - loss: 4.3159e-05
Epoch 22/100
53/53 - 0s - 6ms/step - loss: 4.0293e-05
Epoch 23/100
53/53 - 0s - 6ms/step - loss: 3.6527e-05
Epoch 24/100
53/53 - 0s - 4ms/step - loss: 3.3401e-05
Epoch 25/100
53/53 - 0s - 3ms/step - loss: 3.0053e-05
Epoch 26/100
53/53 - 0s - 4ms/step - loss: 2.7285e-05
Epoch 27/100
53/53 - 0s - 5ms/step - loss: 2.4824e-05
Epoch 28/100
53/53 - 0s - 4ms/step - loss: 2.2033e-05
Epoch 29/100
53/53 - 0s - 3ms/step - loss: 2.0058e-05
Epoch 30/100
53/53 - 0s - 3ms/step - loss: 1.9003e-05
Epoch 31/100
```

```
53/53 - 0s - 6ms/step - loss: 1.8817e-05
Epoch 32/100
53/53 - 0s - 3ms/step - loss: 1.5301e-05
Epoch 33/100
53/53 - 0s - 3ms/step - loss: 1.4457e-05
Epoch 34/100
53/53 - 0s - 6ms/step - loss: 1.2688e-05
Epoch 35/100
53/53 - 0s - 3ms/step - loss: 1.1759e-05
Epoch 36/100
53/53 - 0s - 6ms/step - loss: 1.0354e-05
Epoch 37/100
53/53 - 0s - 6ms/step - loss: 1.0001e-05
Epoch 38/100
53/53 - 0s - 6ms/step - loss: 8.6695e-06
Epoch 39/100
53/53 - 0s - 3ms/step - loss: 8.0949e-06
Epoch 40/100
53/53 - 0s - 6ms/step - loss: 7.3621e-06
Epoch 41/100
53/53 - 0s - 3ms/step - loss: 6.9029e-06
Epoch 42/100
53/53 - 0s - 6ms/step - loss: 6.3113e-06
Epoch 43/100
53/53 - 0s - 4ms/step - loss: 5.8763e-06
Epoch 44/100
53/53 - 0s - 6ms/step - loss: 5.5722e-06
Epoch 45/100
53/53 - 0s - 3ms/step - loss: 5.4152e-06
Epoch 46/100
53/53 - 0s - 3ms/step - loss: 5.4211e-06
Epoch 47/100
53/53 - 0s - 6ms/step - loss: 4.8159e-06
Epoch 48/100
53/53 - 0s - 4ms/step - loss: 4.6382e-06
Epoch 49/100
53/53 - 0s - 5ms/step - loss: 4.5732e-06
Epoch 50/100
53/53 - 0s - 3ms/step - loss: 4.4043e-06
Epoch 51/100
53/53 - 0s - 4ms/step - loss: 4.1975e-06
Epoch 52/100
53/53 - 0s - 6ms/step - loss: 5.0748e-06
Epoch 53/100
53/53 - 0s - 6ms/step - loss: 4.4946e-06
Epoch 54/100
53/53 - 0s - 3ms/step - loss: 4.6110e-06
Epoch 55/100
```

```
53/53 - 0s - 4ms/step - loss: 4.4866e-06
Epoch 56/100
53/53 - 0s - 5ms/step - loss: 4.4330e-06
Epoch 57/100
53/53 - 0s - 6ms/step - loss: 4.1666e-06
Epoch 58/100
53/53 - 0s - 4ms/step - loss: 3.9173e-06
Epoch 59/100
53/53 - 0s - 4ms/step - loss: 3.7772e-06
Epoch 60/100
53/53 - 0s - 6ms/step - loss: 3.6888e-06
Epoch 61/100
53/53 - 0s - 6ms/step - loss: 3.9223e-06
Epoch 62/100
53/53 - 0s - 7ms/step - loss: 4.4327e-06
Epoch 63/100
53/53 - 1s - 11ms/step - loss: 3.9914e-06
Epoch 64/100
53/53 - 0s - 6ms/step - loss: 3.9764e-06
Epoch 65/100
53/53 - 0s - 6ms/step - loss: 3.7522e-06
Epoch 66/100
53/53 - 1s - 11ms/step - loss: 4.2680e-06
Epoch 67/100
53/53 - 0s - 6ms/step - loss: 3.9774e-06
Epoch 68/100
53/53 - 0s - 6ms/step - loss: 3.8383e-06
Epoch 69/100
53/53 - 0s - 4ms/step - loss: 3.7570e-06
Epoch 70/100
53/53 - 0s - 3ms/step - loss: 4.4809e-06
Epoch 71/100
53/53 - 0s - 6ms/step - loss: 3.8895e-06
Epoch 72/100
53/53 - 0s - 3ms/step - loss: 3.9927e-06
Epoch 73/100
53/53 - 0s - 4ms/step - loss: 3.9963e-06
Epoch 74/100
53/53 - 0s - 5ms/step - loss: 3.5958e-06
Epoch 75/100
53/53 - 0s - 4ms/step - loss: 4.3787e-06
Epoch 76/100
53/53 - 0s - 3ms/step - loss: 4.2020e-06
Epoch 77/100
53/53 - 0s - 3ms/step - loss: 3.7460e-06
Epoch 78/100
53/53 - 0s - 4ms/step - loss: 3.5453e-06
Epoch 79/100
```

```
53/53 - 0s - 6ms/step - loss: 4.0586e-06
Epoch 80/100
53/53 - 0s - 3ms/step - loss: 3.6704e-06
Epoch 81/100
53/53 - 0s - 3ms/step - loss: 3.2591e-06
Epoch 82/100
53/53 - 0s - 3ms/step - loss: 3.6940e-06
Epoch 83/100
53/53 - 0s - 3ms/step - loss: 4.1487e-06
Epoch 84/100
53/53 - 0s - 6ms/step - loss: 4.8132e-06
Epoch 85/100
53/53 - 0s - 4ms/step - loss: 4.5441e-06
Epoch 86/100
53/53 - 0s - 3ms/step - loss: 3.9883e-06
Epoch 87/100
53/53 - 0s - 4ms/step - loss: 5.7815e-06
Epoch 88/100
53/53 - 0s - 3ms/step - loss: 4.1191e-06
Epoch 89/100
53/53 - 0s - 6ms/step - loss: 5.2827e-06
Epoch 90/100
53/53 - 0s - 6ms/step - loss: 4.4055e-06
Epoch 91/100
53/53 - 0s - 3ms/step - loss: 4.4897e-06
Epoch 92/100
53/53 - 0s - 4ms/step - loss: 3.9154e-06
Epoch 93/100
53/53 - 0s - 3ms/step - loss: 4.8346e-06
Epoch 94/100
53/53 - 0s - 6ms/step - loss: 3.3952e-06
Epoch 95/100
53/53 - 0s - 3ms/step - loss: 4.3560e-06
Epoch 96/100
53/53 - 0s - 6ms/step - loss: 3.4845e-06
Epoch 97/100
53/53 - 0s - 5ms/step - loss: 3.0818e-06
Epoch 98/100
53/53 - 0s - 4ms/step - loss: 3.6802e-06
Epoch 99/100
53/53 - 0s - 6ms/step - loss: 4.0258e-06
Epoch 100/100
53/53 - 0s - 4ms/step - loss: 4.2251e-06
53/53            1s 7ms/step
6/6            0s 3ms/step
Train Score: 0.12 RMSE
Test Score: 0.83 RMSE
```



[63]:
```
# prompt: calculate rmse rma r2 etc for the above model

# Assuming 'testY' and 'testPredict' are already defined from the previous code

# Calculate RMSE (Root Mean Squared Error)
rmse = np.sqrt(mean_squared_error(testY, testPredict))

# Calculate MSE (Mean Squared Error)
mse = mean_squared_error(testY, testPredict)

# Calculate MAE (Mean Absolute Error)
mae = mean_absolute_error(testY, testPredict)

# Calculate R-squared
r2 = r2_score(testY, testPredict)

print(f'RMSE: {rmse}')
print(f'MSE: {mse}')
print(f'MAE: {mae}')
print(f'R-squared: {r2}')

# You can also calculate other metrics like:
# - MAPE (Mean Absolute Percentage Error)
# - Adjusted R-squared (for multiple regression)
```

```
RMSE: 0.8338649964117724
MSE: 0.6953308322408052
```

```
MAE: 0.7710473469665786
R-squared: 0.9392593838199726
```

[64]:
```python
# prompt: calculate rmse for train and test

# Assuming 'trainY', 'trainPredict', 'testY', and 'testPredict' are already␣
↪defined

# Calculate the training and testing RMSE
train_rmse = np.sqrt(mean_squared_error(trainY, trainPredict))
test_rmse = np.sqrt(mean_squared_error(testY, testPredict))

print(f"Training RMSE: {train_rmse}")
print(f"Testing RMSE: {test_rmse}")
```

```
Training RMSE: 0.11575029663638245
Testing RMSE: 0.8338649964117724
```

[65]:
```python
# prompt: check for overfitting

# Assuming 'trainY', 'trainPredict', 'testY', and 'testPredict' are already␣
↪defined

# Calculate the training and testing RMSE
train_rmse = np.sqrt(mean_squared_error(trainY, trainPredict))
test_rmse = np.sqrt(mean_squared_error(testY, testPredict))

print(f"Training RMSE: {train_rmse}")
print(f"Testing RMSE: {test_rmse}")

# Define a threshold for significance
some_threshold = 5  # You can adjust this value based on your data and model

# Check for overfitting by comparing training and testing RMSE
if test_rmse > train_rmse and (test_rmse - train_rmse) > some_threshold:
  print("Warning: The model might be overfitting.")
  print("The testing RMSE is considerably higher than the training RMSE,␣
↪indicating the model is performing poorly on unseen data.")
else:
  print("The model doesn't appear to be overfitting significantly.")


# Check for underfitting by comparing training and testing RMSE and the␣
↪baseline RMSE
# Create an array of baseline predictions with the same length as test_data
baseline_predictions = np.repeat(np.mean(trainY), len(testY))  # Repeat the␣
↪mean for each test data point
```

```
baseline_rmse = np.sqrt(mean_squared_error(testY, baseline_predictions))  #␣
↪Calculate RMSE using baseline predictions
print(f"Baseline RMSE: {baseline_rmse}")

if train_rmse > baseline_rmse and test_rmse > baseline_rmse:
    print("Warning: The model might be underfitting.")
    print("Both training and testing RMSE are higher than the baseline RMSE,␣
↪indicating the model is not learning effectively.")
elif train_rmse < baseline_rmse and test_rmse > baseline_rmse:
    print("The model is performing better than the baseline on the training␣
↪data but not on the testing data.")
    print("This might indicate that it's not generalizing well or that the␣
↪training data is not representative enough.")
else:
    print("The model doesn't appear to be underfitting significantly.")
```

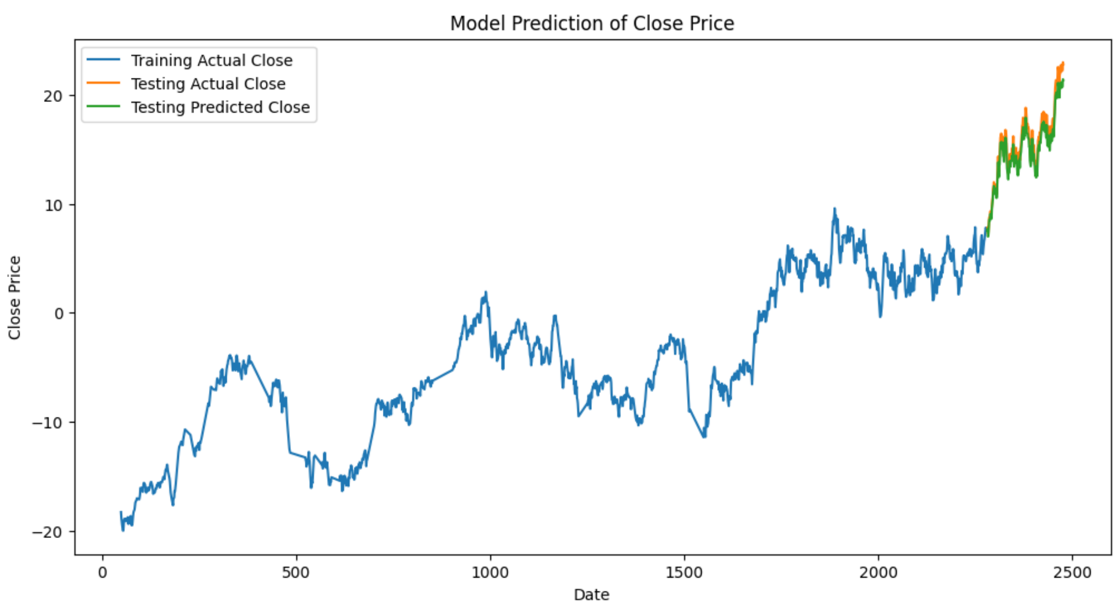
```
Training RMSE: 0.11575029663638245
Testing RMSE: 0.8338649964117724
The model doesn't appear to be overfitting significantly.
Baseline RMSE: 19.819334804725994
The model doesn't appear to be underfitting significantly.
```

LINEAR REGRESSION

[80]:
```
# prompt: write a code to apply Linear regression on NSEI dataset to predict␣
↪close using all attributes. use min max scalar for pre-processing on all␣
↪numeric attributes. use 90% training data and 10% testing data. add all␣
↪required libraries

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

# Assuming 'NSEI' DataFrame is already loaded and prepared
# Extract relevant features for prediction (all attributes except 'Date')
X = NSEI.drop('Date', axis=1)
y = NSEI['Close']

# Normalize the data using MinMaxScaler for numeric attributes
scaler = MinMaxScaler()
numeric_cols = X.select_dtypes(include=np.number).columns
X[numeric_cols] = scaler.fit_transform(X[numeric_cols])

# Split the data into training and testing sets (90% train, 10% test)
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.10,␣
↪random_state=42)

# Create and fit the Linear Regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_train_pred = model.predict(X_train)
y_test_pred = model.predict(X_test)
```

[81]:
```
# prompt: generate graph for predicted test values and training values and␣
↪actual values against time

plt.figure(figsize=(12, 6))
plt.plot(NSEI.index[:train_size], trainY, label='Training Actual Close')
plt.plot(NSEI.index[train_size:], testY, label='Testing Actual Close')
plt.plot(NSEI.index[train_size:], testPredict, label='Testing Predicted Close')
plt.title('Model Prediction of Close Price')
plt.xlabel('Date')
plt.ylabel('Close Price')
plt.legend()
plt.show()
```



[82]:
```
# prompt: calculate rmse rma r2 etc for the above model

# Assuming 'y_test' and 'y_test_pred' are already defined from the previous code
```

```python
# Calculate RMSE (Root Mean Squared Error)
rmse = np.sqrt(mean_squared_error(y_test, y_test_pred))

# Calculate MSE (Mean Squared Error)
mse = mean_squared_error(y_test, y_test_pred)

# Calculate MAE (Mean Absolute Error)
mae = mean_absolute_error(y_test, y_test_pred)

# Calculate R-squared
r2 = r2_score(y_test, y_test_pred)

print(f'RMSE: {rmse}')
print(f'MSE: {mse}')
print(f'MAE: {mae}')
print(f'R-squared: {r2}')

# Calculate the training and testing RMSE
train_rmse = np.sqrt(mean_squared_error(y_train, y_train_pred))
test_rmse = np.sqrt(mean_squared_error(y_test, y_test_pred))

print(f"Training RMSE: {train_rmse}")
print(f"Testing RMSE: {test_rmse}")

# Define a threshold for significance
some_threshold = 5  # You can adjust this value based on your data and model

# Check for overfitting by comparing training and testing RMSE
if test_rmse > train_rmse and (test_rmse - train_rmse) > some_threshold:
  print("Warning: The model might be overfitting.")
  print("The testing RMSE is considerably higher than the training RMSE,␣
 ↪indicating the model is performing poorly on unseen data.")
else:
  print("The model doesn't appear to be overfitting significantly.")


# Check for underfitting by comparing training and testing RMSE and the␣
 ↪baseline RMSE
# Create an array of baseline predictions with the same length as test_data
baseline_predictions = np.repeat(np.mean(y_train), len(y_test))  # Repeat the␣
 ↪mean for each test data point
baseline_rmse = np.sqrt(mean_squared_error(y_test, baseline_predictions))  #␣
 ↪Calculate RMSE using baseline predictions
print(f"Baseline RMSE: {baseline_rmse}")

if train_rmse > baseline_rmse and test_rmse > baseline_rmse:
```

```
    print("Warning: The model might be underfitting.")
    print("Both training and testing RMSE are higher than the baseline RMSE,␣
↪indicating the model is not learning effectively.")
elif train_rmse < baseline_rmse and test_rmse > baseline_rmse:
    print("The model is performing better than the baseline on the training␣
↪data but not on the testing data.")
    print("This might indicate that it's not generalizing well or that the␣
↪training data is not representative enough.")
else:
    print("The model doesn't appear to be underfitting significantly.")
```

```
RMSE: 3.4447020024272506e-13
MSE: 1.186597188552631e-25
MAE: 2.6244727431479446e-13
R-squared: 1.0
Training RMSE: 3.5017712579660177e-13
Testing RMSE: 3.4447020024272506e-13
The model doesn't appear to be overfitting significantly.
Baseline RMSE: 235.88922583473584
The model doesn't appear to be underfitting significantly.
```

RANDOM FORREST

[98]:
```
from sklearn.preprocessing import MinMaxScaler

# Function to preprocess data
def preprocess_data(df):
    # Use only the 'Close' price for prediction
    NSEI = df[['Close']]

    # Initialize the MinMaxScaler to normalize the data
    scaler = MinMaxScaler(feature_range=(0, 1))

    # Scale the 'Close' price data
    scaled_data = scaler.fit_transform(NSEI)

    return scaled_data, scaler
```

[99]:
```
nsei_data, nsei_scaler = preprocess_data(NSEI)
```

[100]:
```
import numpy as np
def create_dataset(data):
    X, y = [], []
    # Loop through the dataset, using each point as input and the next point as␣
↪the target
    for i in range(len(data) - 1):
        X.append(data[i, 0])  # Current day's value
        y.append(data[i + 1, 0])  # Next day's value as target
```

```
    return np.array(X).reshape(-1, 1), np.array(y).reshape(-1, 1)


nsei_X, nsei_y = create_dataset(nsei_data)
```

[101]:
```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import GridSearchCV

# Define the parameter grid for grid search
param_grid = {
    'n_estimators': [100, 200, 300],
    'max_features': ['auto', 'sqrt', 'log2'],
    'max_depth': [10, 20, 30, None],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}

# Function to perform grid search
def perform_grid_search(X, y):
    model = RandomForestRegressor(random_state=42)
    grid_search = GridSearchCV(estimator=model, param_grid=param_grid, cv=5,␣
 ↪n_jobs=-1, verbose=2)
    grid_search.fit(X, y)
    return grid_search.best_estimator_


nsei_rf_best = perform_grid_search(nsei_X, nsei_y)
```

```
Fitting 5 folds for each of 324 candidates, totalling 1620 fits
```

```
/usr/local/lib/python3.10/dist-
packages/joblib/externals/loky/process_executor.py:752: UserWarning: A worker
stopped while some jobs were given to the executor. This can be caused by a too
short worker timeout or by a memory leak.
  warnings.warn(
/usr/local/lib/python3.10/dist-
packages/sklearn/model_selection/_validation.py:540: FitFailedWarning:
540 fits failed out of a total of 1620.
The score on these train-test partitions for these parameters will be set to
nan.
If these failures are not expected, you can try to debug them by setting
error_score='raise'.

Below are more details about the failures:
--------------------------------------------------------------------------------
540 fits failed with the following error:
Traceback (most recent call last):
  File "/usr/local/lib/python3.10/dist-
```

```
packages/sklearn/model_selection/_validation.py", line 888, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "/usr/local/lib/python3.10/dist-packages/sklearn/base.py", line 1466, in
wrapper
    estimator._validate_params()
  File "/usr/local/lib/python3.10/dist-packages/sklearn/base.py", line 666, in
_validate_params
    validate_parameter_constraints(
  File "/usr/local/lib/python3.10/dist-
packages/sklearn/utils/_param_validation.py", line 95, in
validate_parameter_constraints
    raise InvalidParameterError(
sklearn.utils._param_validation.InvalidParameterError: The 'max_features'
parameter of RandomForestRegressor must be an int in the range [1, inf), a float
in the range (0.0, 1.0], a str among {'sqrt', 'log2'} or None. Got 'auto'
instead.

  warnings.warn(some_fits_failed_message, FitFailedWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/model_selection/_search.py:1103:
UserWarning: One or more of the test scores are non-finite: [       nan
nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan 0.63740989 0.63960928 0.63824495
 0.6369774  0.63908423 0.63748633 0.64451814 0.64431897 0.64338895
 0.642795   0.64076445 0.64045438 0.64369228 0.64257665 0.6425135
 0.6446561  0.6446251  0.64390662 0.64629198 0.64506985 0.64479287
 0.64629198 0.64506985 0.64479287 0.64216941 0.64137233 0.64094402
 0.63740989 0.63960928 0.63824495 0.6369774  0.63908423 0.63748633
 0.64451814 0.64431897 0.64338895 0.642795   0.64076445 0.64045438
 0.64369228 0.64257665 0.6425135  0.6446561  0.6446251  0.64390662
 0.64629198 0.64506985 0.64479287 0.64629198 0.64506985 0.64479287
 0.64216941 0.64137233 0.64094402        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
 0.63668213 0.63892726 0.63754874 0.63673118 0.63884434 0.637224
 0.64448475 0.64429047 0.64334586 0.64270013 0.64066945 0.64035454
 0.64361556 0.64249527 0.64242945 0.64464441 0.64460926 0.64389004
 0.64629045 0.64506758 0.64478956 0.64629045 0.64506758 0.64478956
 0.64216754 0.64137128 0.64094225 0.63668213 0.63892726 0.63754874
 0.63673118 0.63884434 0.637224   0.64448475 0.64429047 0.64334586
 0.64270013 0.64066945 0.64035454 0.64361556 0.64249527 0.64242945
 0.64464441 0.64460926 0.64389004 0.64629045 0.64506758 0.64478956
 0.64629045 0.64506758 0.64478956 0.64216754 0.64137128 0.64094225
        nan        nan        nan        nan        nan        nan
```

```
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan 0.6366821  0.63892725 0.63754873
 0.63673118 0.63884434 0.637224   0.64448475 0.64429047 0.64334586
 0.64270013 0.64066945 0.64035454 0.64361556 0.64249527 0.64242945
 0.64464441 0.64460926 0.64389004 0.64629045 0.64506758 0.64478956
 0.64629045 0.64506758 0.64478956 0.64216754 0.64137128 0.64094225
 0.6366821  0.63892725 0.63754873 0.63673118 0.63884434 0.637224
 0.64448475 0.64429047 0.64334586 0.64270013 0.64066945 0.64035454
 0.64361556 0.64249527 0.64242945 0.64464441 0.64460926 0.64389004
 0.64629045 0.64506758 0.64478956 0.64629045 0.64506758 0.64478956
 0.64216754 0.64137128 0.64094225        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
        nan        nan        nan        nan        nan        nan
 0.6366821  0.63892725 0.63754873 0.63673118 0.63884434 0.637224
 0.64448475 0.64429047 0.64334586 0.64270013 0.64066945 0.64035454
 0.64361556 0.64249527 0.64242945 0.64464441 0.64460926 0.64389004
 0.64629045 0.64506758 0.64478956 0.64629045 0.64506758 0.64478956
 0.64216754 0.64137128 0.64094225 0.6366821  0.63892725 0.63754873
 0.63673118 0.63884434 0.637224   0.64448475 0.64429047 0.64334586
 0.64270013 0.64066945 0.64035454 0.64361556 0.64249527 0.64242945
 0.64464441 0.64460926 0.64389004 0.64629045 0.64506758 0.64478956
 0.64629045 0.64506758 0.64478956 0.64216754 0.64137128 0.64094225]
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:1473:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples,), for example using
ravel().
  return fit_method(estimator, *args, **kwargs)
```

[102]:

```python
import numpy as np
from sklearn.metrics import r2_score, mean_squared_error

# Define evaluation metrics
def rmse(y_true, y_pred):
    return np.sqrt(np.mean((y_pred - y_true) ** 2))

def mape(y_true, y_pred):
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100

def mbe(y_true, y_pred):
    return np.mean(y_pred - y_true)

# Function to evaluate the model
```

```
def evaluate_model(model, X, y, scaler):
    # Predict using the model
    predicted = model.predict(X)

    # Inverse transform the predictions and true values to the original scale
    predicted = scaler.inverse_transform(predicted.reshape(-1, 1))
    y = scaler.inverse_transform(y.reshape(-1, 1))

    # Calculate evaluation metrics
    rmse_val = rmse(y, predicted)
    mape_val = mape(y, predicted)
    mbe_val = mbe(y, predicted)
    mse_val = mean_squared_error(y, predicted)
    r2_val = r2_score(y, predicted)

    # Print metrics
    print(f"Evaluation Metrics:")
    print(f"RMSE: {rmse_val}")
    print(f"MAPE: {mape_val}")
    print(f"MBE: {mbe_val}")
    print(f"MSE: {mse_val}")
    print(f"R²: {r2_val}")

    return rmse_val, mape_val, mbe_val, mse_val, r2_val
```

[103]:
```
nsei_rmse, nsei_mape, nsei_mbe, nsei_mse, nsei_rsquare =␣
↪evaluate_model(nsei_rf_best, nsei_X, nsei_y, nsei_scaler)
```

```
Evaluation Metrics:
RMSE: 13.170333456954882
MAPE: 0.6626018666995912
MBE: -0.027622176481664117
MSE: 173.45768336738513
R²: 0.997066017846212
```

[104]:
```
# prompt: find evaluation matrix r2, rmse, mse, mae

# Assuming you have already trained your model and have nsei_rf_best, nsei_X,␣
↪nsei_y, and nsei_scaler defined as in your provided code.

# You can directly call the evaluate_model function to get the desired metrics.
nsei_rmse, nsei_mape, nsei_mbe, nsei_mse, nsei_rsquare =␣
↪evaluate_model(nsei_rf_best, nsei_X, nsei_y, nsei_scaler)

# Print the individual metrics
print(f"R-squared: {nsei_rsquare}")
print(f"RMSE: {nsei_rmse}")
```

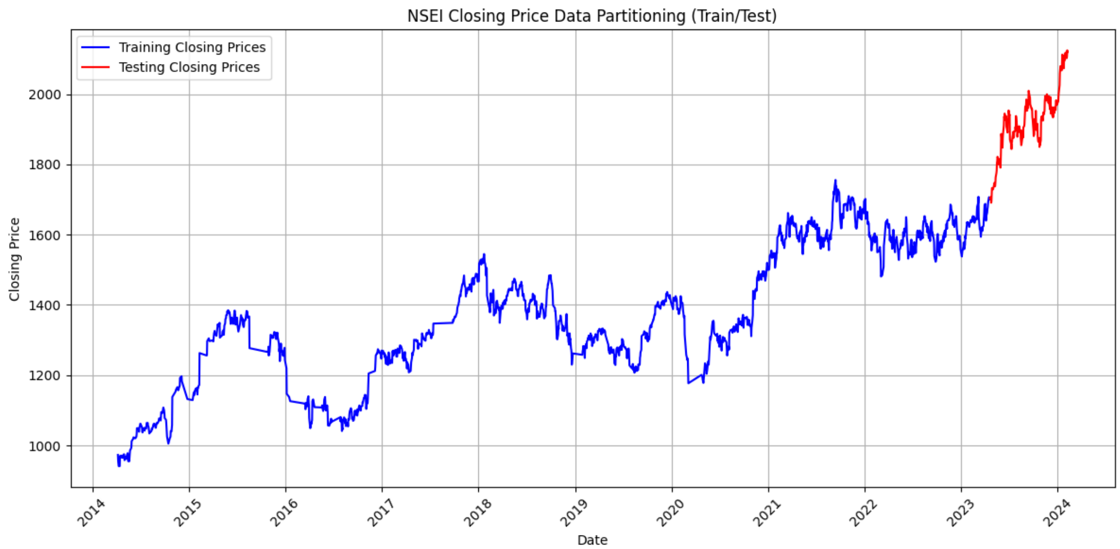
```
print(f"MSE: {nsei_mse}")
print(f"MAE: {nsei_rmse}")  # Assuming you are using RMSE as a proxy for MAE,␣
↪as MAE is not calculated directly in the provided code.
```

```
Evaluation Metrics:
RMSE: 13.170333456954882
MAPE: 0.6626018666995912
MBE: -0.027622176481664117
MSE: 173.45768336738513
R²: 0.997066017846212
R-squared: 0.997066017846212
RMSE: 13.170333456954882
MSE: 173.45768336738513
MAE: 13.170333456954882
```

[105]:

```
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split

# Assuming 'NSEI' DataFrame exists with 'Date' and 'Close' columns
# Also assuming 'nsei_X', 'nsei_y', and 'nsei_scaler' are defined

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(nsei_X, nsei_y, test_size=0.
↪1, random_state=42, shuffle=False)

# Ensure the dates align with the data split
train_dates = NSEI['Date'][:len(y_train)].reset_index(drop=True)
test_dates = NSEI['Date'][len(y_train):len(y_train) + len(y_test)].
↪reset_index(drop=True)

# Plotting
plt.figure(figsize=(12, 6))

# Plot training data
plt.plot(train_dates, nsei_scaler.inverse_transform(y_train.reshape(-1, 1)),␣
↪label='Training Closing Prices', color='blue')

# Plot testing data
plt.plot(test_dates, nsei_scaler.inverse_transform(y_test.reshape(-1, 1)),␣
↪label='Testing Closing Prices', color='red')

# Set labels, title, legend, and grid
plt.xlabel('Date')
plt.ylabel('Closing Price')
plt.title('NSEI Closing Price Data Partitioning (Train/Test)')
plt.legend()
plt.grid(True)
```

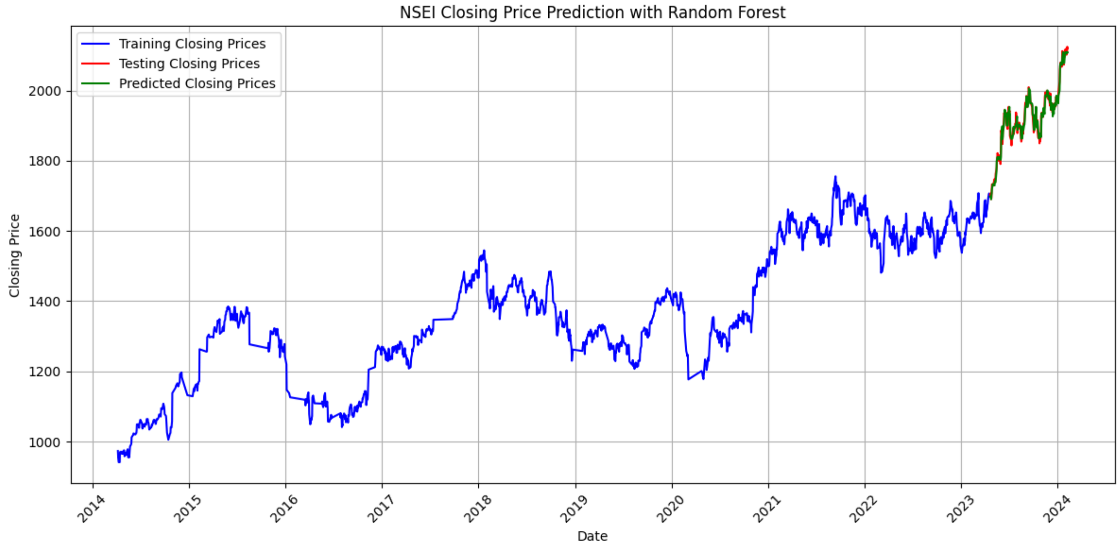
```
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.tight_layout()
plt.show()
```



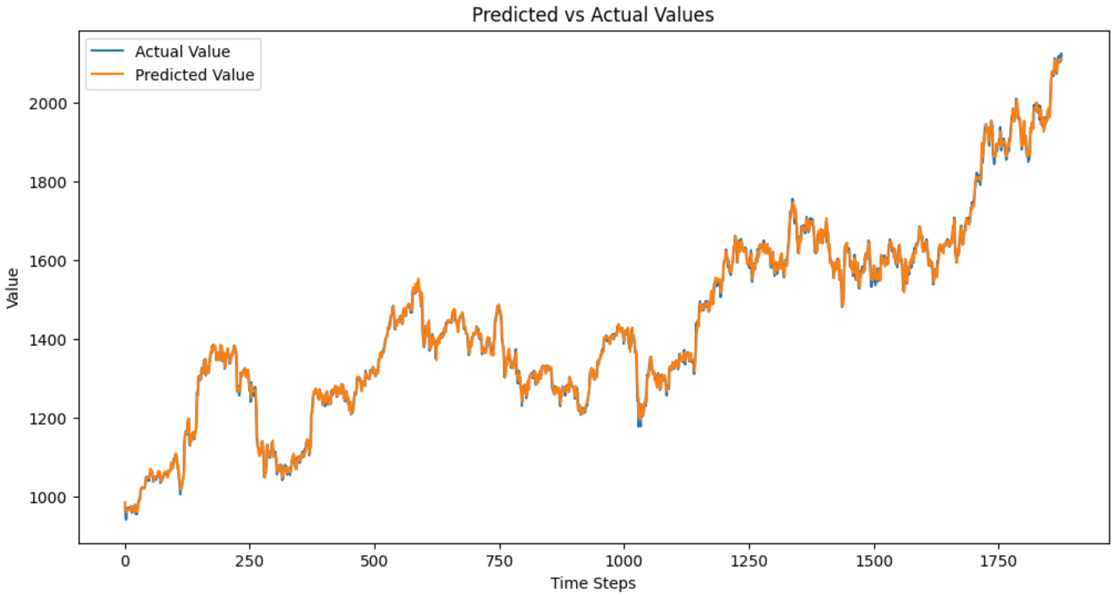
[106]:
```
# prompt: generate graph for training data, testing data and predicted data

# Assuming you have already trained your model and have nsei_rf_best, nsei_X,␣
↪nsei_y, and nsei_scaler defined as in your provided code.

# You can directly call the evaluate_model function to get the desired metrics.
nsei_rmse, nsei_mape, nsei_mbe, nsei_mse, nsei_rsquare =␣
↪evaluate_model(nsei_rf_best, nsei_X, nsei_y, nsei_scaler)

# Print the individual metrics
print(f"R-squared: {nsei_rsquare}")
print(f"RMSE: {nsei_rmse}")
print(f"MSE: {nsei_mse}")
print(f"MAE: {nsei_rmse}")  # Assuming you are using RMSE as a proxy for MAE,␣
↪as MAE is not calculated directly in the provided code.

# Assuming 'NSEI' DataFrame exists with 'Date' and 'Close' columns
# Also assuming 'nsei_X', 'nsei_y', and 'nsei_scaler' are defined

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(nsei_X, nsei_y, test_size=0.
↪1, random_state=42, shuffle=False)

# Predict on the test set
```

```
y_pred = nsei_rf_best.predict(X_test)

# Ensure the dates align with the data split
train_dates = NSEI['Date'][:len(y_train)].reset_index(drop=True)
test_dates = NSEI['Date'][len(y_train):len(y_train) + len(y_test)].
    reset_index(drop=True)

# Inverse transform the predictions and true values to the original scale
y_pred_original = nsei_scaler.inverse_transform(y_pred.reshape(-1, 1))
y_test_original = nsei_scaler.inverse_transform(y_test.reshape(-1, 1))

# Plotting
plt.figure(figsize=(12, 6))

# Plot training data
plt.plot(train_dates, nsei_scaler.inverse_transform(y_train.reshape(-1, 1)),
    label='Training Closing Prices', color='blue')

# Plot testing data
plt.plot(test_dates, y_test_original, label='Testing Closing Prices',
    color='red')

# Plot predicted data
plt.plot(test_dates, y_pred_original, label='Predicted Closing Prices',
    color='green')

# Set labels, title, legend, and grid
plt.xlabel('Date')
plt.ylabel('Closing Price')
plt.title('NSEI Closing Price Prediction with Random Forest')
plt.legend()
plt.grid(True)
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.tight_layout()
plt.show()
```

```
Evaluation Metrics:
RMSE: 13.170333456954882
MAPE: 0.6626018666995912
MBE: -0.027622176481664117
MSE: 173.45768336738513
R²: 0.997066017846212
R-squared: 0.997066017846212
RMSE: 13.170333456954882
MSE: 173.45768336738513
MAE: 13.170333456954882
```



[107]:
```python
import matplotlib.pyplot as plt
import seaborn as sns

# Predicted vs. Actual values plot
def plot_predicted_vs_actual(y_true, y_pred, title):
    plt.figure(figsize=(12, 6))
    plt.plot(y_true, label="Actual Value")
    plt.plot(y_pred, label="Predicted Value")
    plt.title('Predicted vs Actual Values')
    plt.xlabel('Time Steps')
    plt.ylabel('Value')
    plt.legend()
    plt.show()


plot_predicted_vs_actual(
    nsei_scaler.inverse_transform(nsei_y.reshape(-1, 1)),
    nsei_scaler.inverse_transform(nsei_rf_best.predict(nsei_X).reshape(-1, 1)),
    "NIFTY: Predicted vs Actual"
)
```



[108]:
```
# prompt: check for underfitting using rmse of train and test data

# Calculate RMSE for training data
y_train_pred = nsei_rf_best.predict(X_train)
train_rmse = rmse(y_train, y_train_pred)

# Calculate RMSE for testing data
y_test_pred = nsei_rf_best.predict(X_test)
test_rmse = rmse(y_test, y_test_pred)

print(f"Training RMSE: {train_rmse}")
print(f"Testing RMSE: {test_rmse}")

# Check for underfitting
if train_rmse > test_rmse:
  print("Possible underfitting detected.")
elif train_rmse == test_rmse:
  print("The model might be very simple or the dataset might be too small.")
else:
  print("The model is likely not underfitting.")
```

```
Training RMSE: 0.22575134934977067
Testing RMSE: 0.11083497597477927
Possible underfitting detected.
```

[109]:
```
# prompt: check for overfitting using rmse of train and test data
```

```python
# Calculate RMSE for training data
y_train_pred = nsei_rf_best.predict(X_train)
train_rmse = rmse(y_train, y_train_pred)

# Calculate RMSE for testing data
y_test_pred = nsei_rf_best.predict(X_test)
test_rmse = rmse(y_test, y_test_pred)

print(f"Training RMSE: {train_rmse}")
print(f"Testing RMSE: {test_rmse}")

# Check for overfitting
if test_rmse > train_rmse:
    print("Possible overfitting detected.")
    print("The model is performing significantly better on the training data␣
 ↪than on the test data.")
else:
    print("The model is likely not overfitting.")
```

```
Training RMSE: 0.22575134934977067
Testing RMSE: 0.11083497597477927
The model is likely not overfitting.
```