EX:No.8 221501042

**Create an ARIMA model for time series forecasting.**

**Aim:**

Write a program to create an ARIMA model for time series forecasting.

**Algorithm:**

1. **Import necessary libraries**:  
Import numpy, pandas, matplotlib.pyplot, ARIMA from statsmodels.tsa.arima.model, and mean\_squared\_error from sklearn.metrics.

2. **Load the dataset**:  
Read the weather dataset using pandas.read\_csv() and parse the 'Date' column as datetime. Set the 'Date' column as the index.

3. **Select the target column**:  
Extract the 'Temperature' column from the DataFrame for time series forecasting.

4. **Split the data**:  
Calculate the training size as 80% of the total data.  
Split the data into training and testing sets using this calculated size.

5. **Define and fit the ARIMA model**:  
Initialize the ARIMA model with training data and set parameters (p=5, d=1, q=0).  
Fit the model using .fit().

6. **Forecast future values**:  
Forecast the temperature values for the length of the test set using .forecast().

7.**Visualize the training data**:  
Create a line plot for the training data using matplotlib.

8.**Visualize the actual test data**:  
Create a line plot for the test (actual) data.

9.**Visualize the forecasted vs actual test data**:  
Plot the ARIMA forecast along with the actual test values on the same graph to compare performance.

10.**Evaluate the model**:  
Calculate the Root Mean Squared Error (RMSE) between the forecast and actual test values using mean\_squared\_error().

11.**Print the RMSE value**.

**Code:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from statsmodels.tsa.arima.model import ARIMA

from sklearn.metrics import mean\_squared\_error

df = pd.read\_csv(r"C:\Users\harsh\Downloads\cleaned\_weather.csv", parse\_dates=['date'], index\_col='date')

time\_series = df['T']

train\_size = int(len(time\_series) \* 0.8)

train, test = time\_series[:train\_size], time\_series[train\_size:]

# Fit ARIMA model (p=5, d=1, q=0) - you can tune these parameters based on AIC/BIC or grid search

model = ARIMA(train, order=(5, 1, 0))

model\_fit = model.fit()

forecast = model\_fit.forecast(steps=len(test))

plt.figure(figsize=(10, 4))

plt.plot(train, label='Training Data', color='blue')

plt.title('Training Data')

plt.xlabel('Date')

plt.ylabel('Temperature')

plt.legend()

plt.grid(True)

plt.tight\_layout()

plt.show()

plt.figure(figsize=(10, 4))

plt.plot(test, label='Actual Test Data', color='green')

plt.title('Actual Test Data')

plt.xlabel('Date')

plt.ylabel('Temperature')

plt.legend()

plt.grid(True)

plt.tight\_layout()

plt.show()

plt.figure(figsize=(10, 4))

plt.plot(test.index, forecast, label='ARIMA Forecast', color='red')

plt.plot(test, label='Actual Test Data', color='green')

plt.title('ARIMA Forecast vs Actual')

plt.xlabel('Date')

plt.ylabel('Temperature')

plt.legend()

plt.grid(True)

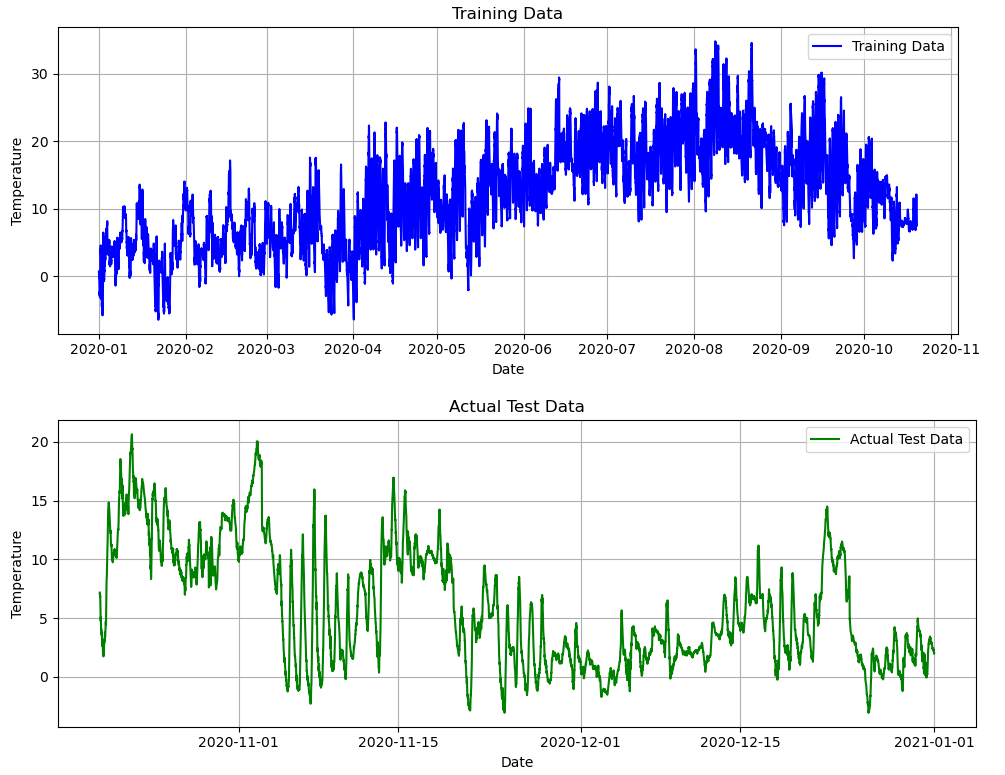
plt.tight\_layout()

plt.show()

rmse = np.sqrt(mean\_squared\_error(test, forecast))

print(f'RMSE: {rmse:.2f}')

**Output:**

****

**A graph showing a green line

AI-generated content may be incorrect.**

**Result:**

Thus, the program to create an ARIMA model for time series forecasting was created successfully.