EX:No.4 221501043

25/01/25

**Program to Develop a linear regression model for forecasting time series data.**

**Aim:**

Write a program to implement time series data for import library, load data, Preprocessing and visualising.

**Algorithm:**

1. **Load the Data**:
   * Read the CSV file containing the weather data.
   * Parse the date column as a datetime index.
2. **Clean the Data**:
   * Handle missing values by performing forward and backward filling.
   * Drop any remaining NaN values.
3. **Normalize the Data**:
   * Apply **Min-Max Scaling** to normalize each column's values between 0 and 1.
4. **Add Time-Based Features**:
   * Extract additional features from the datetime index: day, month and year
5. **Visualize the Data**:
   * Plot the time series for a specific column (e.g., temperature T) over time.
6. **Execute the Program**:
   * Sequentially call the functions to load, clean, normalize, add features, and visualize the data.

**Code:**

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

from sklearn.linear\_model import LinearRegression

# Load the dataset

file\_path = "/mnt/data/daily-minimum-temperatures-in-me.csv"

df = pd.read\_csv(file\_path)

# Convert 'Date' column to datetime format and set it as index

df['Date'] = pd.to\_datetime(df['Date'])

df.set\_index('Date', inplace=True)

# Rename the temperature column for easier access

df.rename(columns={'Daily minimum temperatures': 'Temperature'}, inplace=True)

# Convert Temperature column to numeric, forcing errors to NaN and then dropping them

df['Temperature'] = pd.to\_numeric(df['Temperature'], errors='coerce')

df.dropna(inplace=True)

# Apply Linear Regression

X = np.array((df.index - df.index.min()).days).reshape(-1, 1) # Convert dates to numerical values

y = df['Temperature'].values

model = LinearRegression()

model.fit(X, y)

df['Trend'] = model.predict(X)

# Plot original data and linear regression trend

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

plt.plot(df.index, df['Temperature'], label='Original Data', alpha=0.5)

plt.plot(df.index, df['Trend'], label='Linear Regression Trend', color='blue', linestyle='dashed')

plt.xlabel('Date')

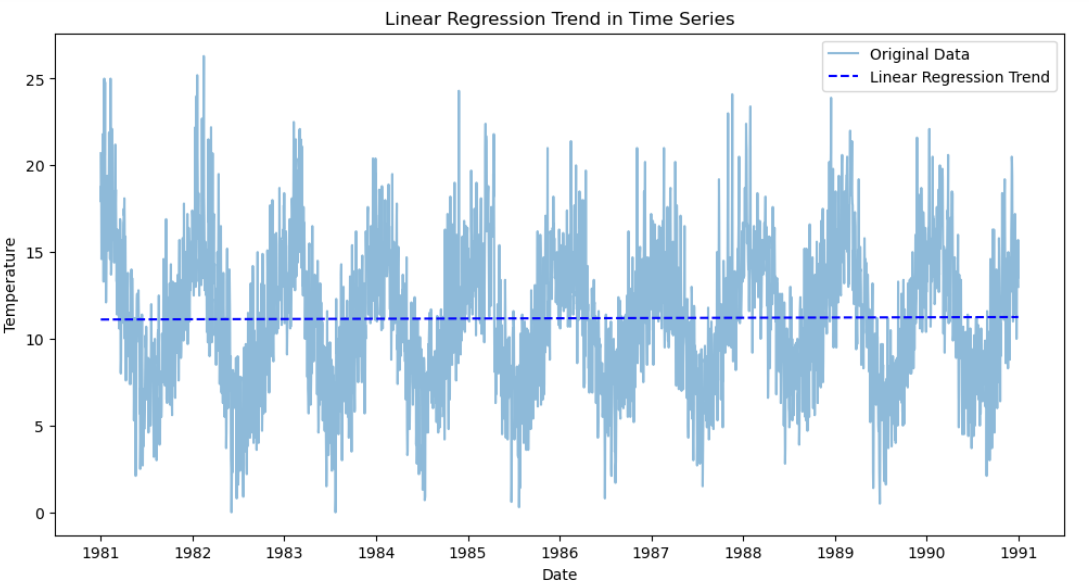
plt.ylabel('Temperature')

plt.title('Linear Regression Trend in Time Series')

plt.legend()

plt.show()

**Output:**



**Result:**

Thus, the program using the time series data implementation has been done successfully.