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

**EX:No.5 DATE:7/02/25**

# AIM:

Develop a linear regression model for forecasting time series data.

## OBJECTIVE:

To develop a linear regression model to forecast future air pollution levels based on historical data.

## BACKGROUND:

* **Linear regression** models the relationship between dependent and independent variables.
* In time-series forecasting, **time** (e.g., year, month) can be an independent variable for predicting pollution levels.
* Linear regression can help predict future pollution trends based on historical data.
* The model is simple but effective for linear relationships and can be used for short-term forecasts.

## SCOPE OF THE PROGRAM:

* Load and clean air pollution data (2012-2021).
* Use **time** (month/year) as a feature for regression.
* Build a **linear regression model** for predicting future pollution levels.
* Evaluate the model performance with metrics like **mean squared error (MSE)**.

**CODE:**

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.linear\_model import LinearRegression from sklearn.model\_selection import train\_test\_split

# Load data

df = pd.read\_csv("/content/us\_air\_pollution\_2012\_2021\_updated.csv") df['Date'] = pd.to\_datetime(df['Date'])

df['Date\_ordinal'] = df['Date'].map(lambda x: x.toordinal()) # Convert Date to numerical

# Features & Target

X = df[['Date\_ordinal']]

y = df["PM2.5 (µg/m³)"] # Update column name if different

# Train-Test Split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, shuffle=False)

# Train Model

model = LinearRegression() model.fit(X\_train, y\_train)

# Predict

y\_pred = model.predict(X\_test)

# Plot plt.figure(figsize=(10, 5))

plt.plot(df['Date'], y, label="Original", color='blue') plt.plot(df.iloc[len(X\_train):]['Date'], y\_pred, label="Forecast", color='red') plt.xlabel("Date")

plt.ylabel("Pollution Level") plt.title("Linear Regression Forecasting") plt.legend()

plt.show()

# OUTPUT:

**RESULT:**

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