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

**EX:No.5**

**DATE:**

# 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

df = pd.read\_csv("your\_dataset.csv", encoding="latin1")

df.columns = df.columns.str.strip()

df['Order Date'] = pd.to\_datetime(df['Order Date'], errors='coerce')

df['Date\_ordinal'] = df['Order Date'].map(lambda x: x.toordinal())

df = df.dropna(subset=['Sales'])

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

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

y = df['Sales']

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

model = LinearRegression()

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

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

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

plt.plot(df['Order Date'], y, label="Original Sales", color='blue', alpha=0.5)

plt.plot(df.iloc[len(X\_train):]['Order Date'], y\_pred, label="Forecasted Sales", color='red')

plt.xlabel("Date", fontsize=12)

plt.ylabel("Sales", fontsize=12)

plt.title("Sales Forecasting using Linear Regression", fontsize=14)

plt.legend()

plt.grid(True, linestyle="--", alpha=0.6)

plt.show()

# OUTPUT:

# 

**RESULT:**

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