**EX: 5 DEVELOP A LINEAR REGRESSION MODEL DATE: FOR FORECASTING TIME SERIES DATA**

**AIM:**

To develop a Linear Regression model for forecasting time series data and evaluate its performance using error metrics and residual analysis.

**ALGORITHM:**

1. Load Data – Import the dataset and convert the date column to datetime format.

2. Preprocess Data – Convert dates into numerical values for regression. 3. Split Data – Divide the dataset into training and testing sets. 4. Train Model – Fit a Linear Regression model using the training data. 5. Make Predictions – Predict values on the test set using the trained model.

6. Evaluate Model – Calculate MAE, MSE, and RMSE for performance assessment.

7. Visualize Results – Plot a residual plot to analyze prediction errors.

**PROGRAM:**

# Install necessary libraries (if not already installed)

!pip install pandas scikit-learn

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

from sklearn.preprocessing import StandardScaler

# 1. Load the dataset from a CSV file (upload the weather.csv file in Colab)

from google.colab import files

uploaded = files.upload()

# Load the data into a pandas DataFrame

df = pd.read\_csv(next(iter(uploaded)))

# 2. Preprocess the data

# Convert categorical columns (e.g., Station.City, Station.State) using one-hot encoding

df = pd.get\_dummies(df, columns=['Station.City', 'Station.State'], drop\_first=True)

# Convert date columns to numeric format (e.g., year, month)

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

df['Year'] = df['Date.Full'].dt.year

df['Month'] = df['Date.Full'].dt.month

df['Day'] = df['Date.Full'].dt.day

# Select features (X) and target (y)

X = df[['Month', 'Day', 'Year', 'Data.Temperature.Avg Temp', 'Data.Temperature.Max Temp', 'Data.Temperature.Min Temp', 'Data.Wind.Direction', 'Data.Wind.Speed'] + [col for col in df.columns if 'Station.City' in col or 'Station.State' in col]]

y = df['Data.Precipitation'] # Target variable is Data.Precipitation

# 3. Split the data into training and test sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# 4. Standardize the features (optional but often useful for regression)

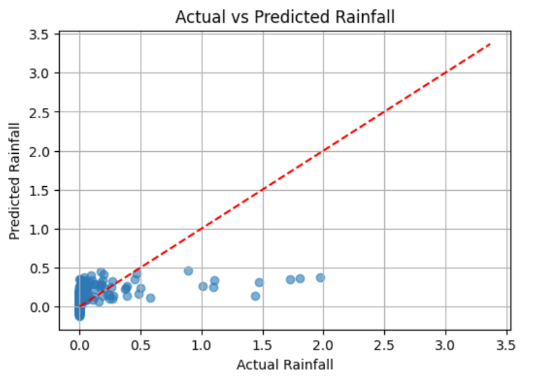
scaler = StandardScaler()

X\_train\_scaled = scaler.fit\_transform(X\_train)

X\_test\_scaled = scaler.transform(X\_test)

# 5. Train the regression model (e.g.,

**OUTPUT:**

****

**RESULT:**

Thus, the Linear Regression model for forecasting Time Series data using Air passenger dataset was successfully developed and evaluated.