**Implement program for linear regression using time series data**

**AIM:**

To implement program for linear regression using time series data using jupyter notebook.

**ALGORITHM:**

1. Load the taxi dataset then clean and load the values

2. Preprocess the values such as variables in the dataset

3. perform the linear regression using the dataset

**CODE:**

import pandas as pd

import matplotlib.pyplot as plt

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

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_absolute\_error

file\_path = "Downloads/train.csv/taxi dataset.csv"

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

df['datetime'] = pd.to\_datetime(df[['year', 'month', 'day']]) + pd.to\_timedelta(df['hour\_of\_day'], unit='h')

df = df.sort\_values(by='datetime')

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

df\_daily = df[['total\_amount']].resample('D').sum()

df\_daily['day'] = df\_daily.index.day

df\_daily['month'] = df\_daily.index.month

df\_daily['year'] = df\_daily.index.year

df\_daily['day\_of\_week'] = df\_daily.index.dayofweek

X = df\_daily[['day', 'month', 'year', 'day\_of\_week']]

y = df\_daily['total\_amount']

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

model = LinearRegression()

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

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

mae = mean\_absolute\_error(y\_test, y\_pred)

print(f"Mean Absolute Error: {mae}")

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

plt.plot(df\_daily.index, df\_daily['total\_amount'], label="Actual", color='blue')

plt.plot(X\_test.index, y\_pred, label="Predicted", color='red', linestyle="dashed")

plt.xlabel("Date")

plt.ylabel("Total Fare Amount")

plt.title("Daily Taxi Fare Amount Prediction using Linear Regression")

plt.legend()

plt.grid()

plt.show()

**OUTPUT:**



**RESULT:**

The program to execute linear regression using time series data completed successfully and the output is verified.