

EX:No.6

DATE:29/03/25

Implement program to apply moving average smoothing for data preparation and time series forecasting

AIM:

Write a program to Implement program to apply moving average smoothing for data preparation and time series forecasting

ALGORITHM:

- Step 1: Install required libraries (if not already installed).
- Step 2: Import necessary libraries (pandas, numpy, matplotlib).
- Step 3: Load air pollution data, parse dates, and set 'date' as the index.
- Step 4: Remove duplicate timestamps and fill missing values.
- Step 5: Select the 'pollution_today' column.
- Step 6: Remove outliers using the IQR method.
- Step 7: Ensure daily data frequency.
- Step 8: Resample to weekly average (optional, not used in the plot).
- Step 9: Create a figure and plot daily pollution levels as a line graph.
- Step 10: Set labels, title, and legend for the plot.
- Step 11: Show the plot.

CODE:

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the time series data
file_path = r"C:\Users\hemhe\Downloads\archive\symbols_valid_meta.csv"
df = pd.read_csv(file_path, parse_dates=["Date"], index_col="Date")
# Verify column
if "Total Accidents" not in df.columns:
    raise ValueError("Column 'Total Accidents' not found.")

ts = df["Total Accidents"].dropna()
```

```

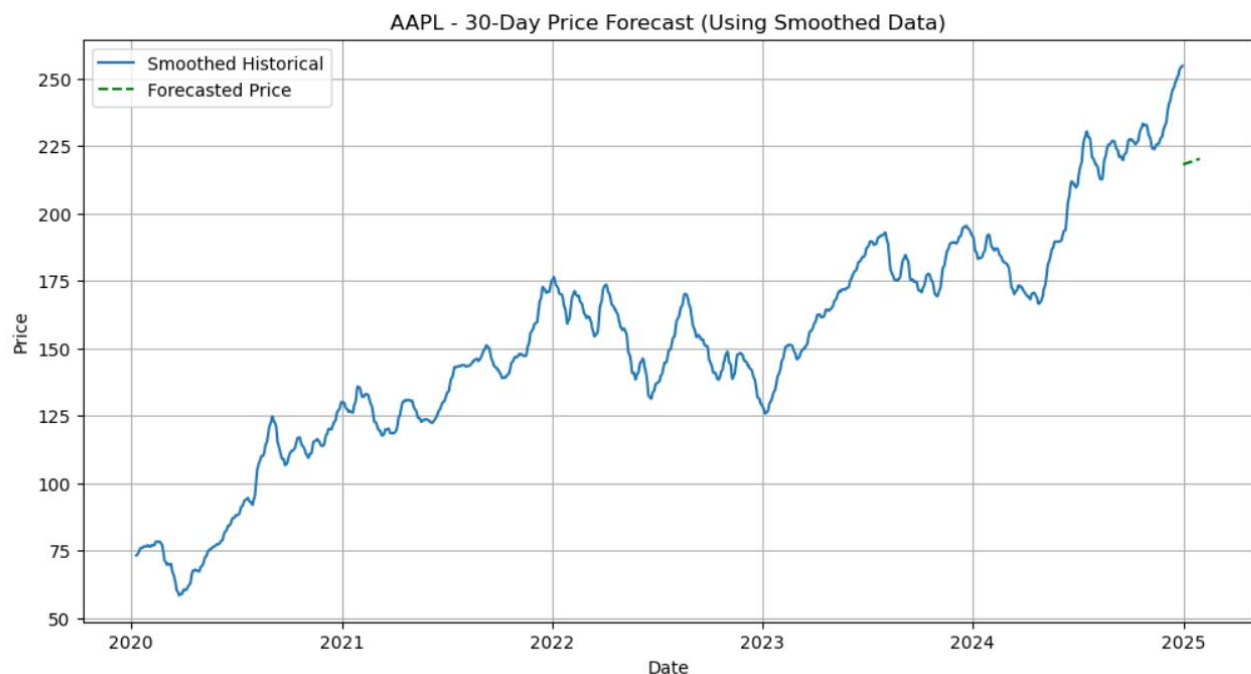
# Apply Moving Average Smoothing
window_size = 12 # 12-month moving average
moving_avg = ts.rolling(window=window_size).mean()

# Forecasting using Moving Average
future_steps = 12 # Forecast for next 12 months
forecast_values = moving_avg[-1] # Last smoothed value as a baseline forecast
future_dates = pd.date_range(start=ts.index[-1], periods=future_steps + 1, freq="M")[1:]
forecast_series = pd.Series([forecast_values] * future_steps, index=future_dates)

# Plot Original, Smoothed & Forecasted Series
plt.figure(figsize=(12, 5))
plt.plot(ts, label="Original Time Series", alpha=0.5)
plt.plot(moving_avg, label=f"{window_size}-month Moving Average", color="red", linestyle="dashed")
plt.plot(forecast_series, label="Forecast (Moving Avg)", color="green", linestyle="dotted")
plt.xlabel("Date")
plt.ylabel("Total Accidents")
plt.title("Moving Average Smoothing & Forecasting")
plt.legend()
plt.grid(True)
plt.show()

```

OUTPUT:



RESULT:

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