**EXP 6:- Implement program to apply moving average smoothing for data preparation and time series forecasting**

**AIM:**

To apply moving average smoothing on the daily confirmed COVID-19 cases over time

**ALGORITHM:**

**Step 1: Load the Dataset**

* Import the necessary libraries (pandas, matplotlib, ARIMA).
* Load the dataset from a CSV file into a pandas DataFrame.

**Step 2: Clean and Preprocess the Data**

* Clean the column names by stripping any leading or trailing spaces.
* Convert the Date column to a datetime format.
* Set the Date column as the index of the DataFrame for time series analysis.
* Sort the data by date to ensure chronological order.

**Step 3: Aggregate Global Daily Confirmed Cases**

* Group the data by Date and sum the confirmed cases for each day.

**Step 4: Create Daily New Cases**

* Calculate the daily new cases by finding the difference in confirmed cases between consecutive days.
* Remove any rows with missing values (such as the first row).

**Step 5: Add Moving Averages (SMA and EMA)**

* Calculate a 5-day Simple Moving Average (SMA) for the daily new cases.
* Calculate a 5-day Exponential Moving Average (EMA) for the daily new cases.

**Step 6: Fit ARIMA Model for Forecasting**

* Fit an ARIMA model to the daily new cases with the appropriate parameters (e.g., p=1, d=1, q=1).
* Use the model to forecast the next 5 days of daily new cases.

**Step 7: Prepare Forecast Data**

* Create a DataFrame to store the forecasted daily new cases and their corresponding dates.
* Combine the actual daily cases and the forecasted values into a single DataFrame.

**Step 8: Plot the Data**

* Plot the original daily new cases, the 5-day SMA, the 5-day EMA, and the forecasted daily cases on a single graph.
* Customize the plot by adding labels, a title, and a legend.

**Step 9: Save the Processed Data**

* Save the combined dataset (including both actual and forecasted daily new cases) to a new CSV file.

**PROGRAM:**

**OUTPUT:**

import pandas as pd

import matplotlib.pyplot as plt

from statsmodels.tsa.arima.model import ARIMA

file\_path = "/time-series-19-covid-combined.csv"

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

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

df["Date"] = pd.to\_datetime(df["Date"], format="%Y-%m-%d")

df.set\_index("Date", inplace=True)

df.sort\_index(inplace=True)

daily\_df = df.groupby(df.index).sum()

daily\_df["daily\_cases"] = daily\_df["Confirmed"].diff().dropna()

daily\_df = daily\_df.dropna(subset=["daily\_cases"])

daily\_df["SMA\_5"] = daily\_df["daily\_cases"].rolling(window=5).mean()

daily\_df["EMA\_5"] = daily\_df["daily\_cases"].ewm(span=5, adjust=False).mean()

model = ARIMA(daily\_df["daily\_cases"], order=(1, 1, 1))

model\_fit = model.fit()

forecast = model\_fit.forecast(steps=5)

future\_dates = pd.date\_range(start=daily\_df.index.max() + pd.Timedelta(days=1), periods=5)

forecast\_df = pd.DataFrame({"Date": future\_dates, "Forecasted\_daily\_cases": forecast})

forecast\_df.set\_index("Date", inplace=True)

combined\_df = pd.concat([daily\_df, forecast\_df], axis=0)

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

plt.plot(daily\_df.index, daily\_df["daily\_cases"], label="Original Daily Cases", linestyle='dotted', marker='o', alpha=0.6)

plt.plot(daily\_df.index, daily\_df["SMA\_5"], label="5-Day SMA", color='red')

plt.plot(daily\_df.index, daily\_df["EMA\_5"], label="5-Day EMA", color='green')

plt.plot(forecast\_df.index, forecast\_df["Forecasted\_daily\_cases"], label="Forecasted Daily Cases", color="blue", marker="x", linestyle="--")

plt.legend()

plt.title("COVID-19 Daily Cases: Smoothing and Forecasting")

plt.xlabel("Date")

plt.ylabel("Daily New Cases")

plt.tight\_layout()

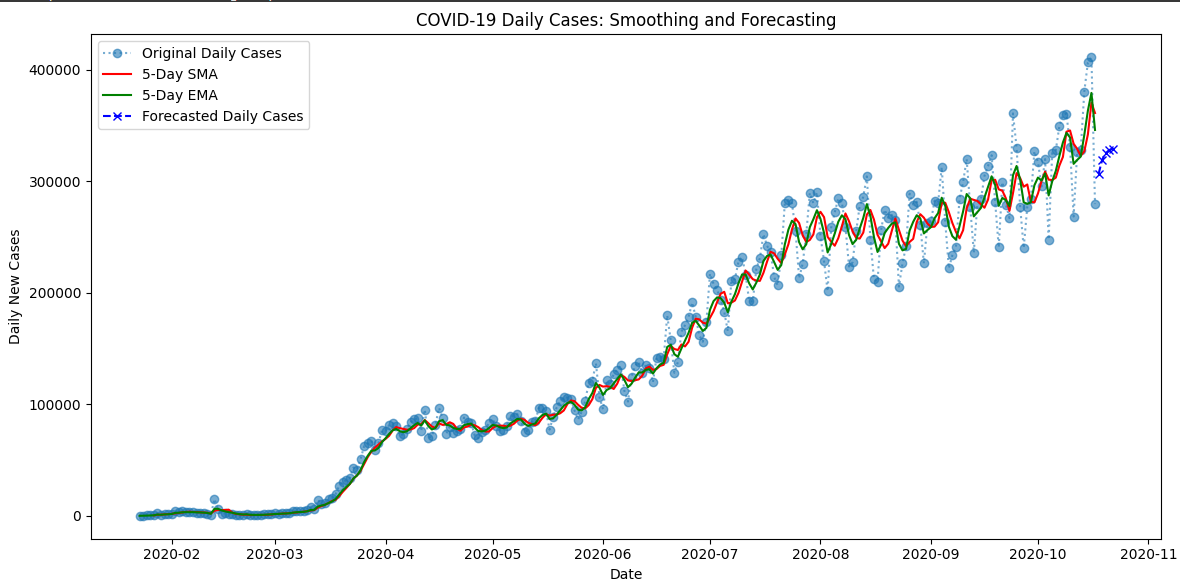
plt.show()

output\_path = "/mnt/data/covid\_smoothed\_forecast.csv"

combined\_df.to\_csv(output\_path)

print(f"✅ Smoothed and forecasted data saved to: {output\_path}")

**OUTPUT:**



**RESULT:**

Moving average smoothing effectively reduces fluctuations in the **RANK** column over **YEAR**, highlighting overall trends. The **SMA** provides a stable long-term trend, while the **EMA** quickly adapts to recent changes, making trend analysis and forecasting more reliable.