**Exp no: 8 Create an ARIMA model for time series forecasting. 221501022**

**Date: 8/4/25**

**Aim:**

The aim of this project is to develop a time series forecasting model using the ARIMA (AutoRegressive Integrated Moving Average) technique to accurately predict future superstores volumes based on historical monthly data.

**Objectives:**

The objectives of this project revolve around understanding and modeling the temporal structure of the data. This includes loading and preprocessing the Autism dataset, visualizing trends and seasonal patterns, and checking for stationarity to ensure the dataset meets the assumptions required for ARIMA modeling. The project also focuses on selecting optimal ARIMA parameters, training the model, validating its performance using forecasting metrics, and ultimately generating forecasts for future time periods.

**Background/Scope:**

Time series forecasting plays a crucial role in sectors where planning and resource allocation are dependent on future demand. The Autism dataset, which records monthly totals of international superstoress from 1949 to 1960, serves as a standard benchmark for time series analysis. ARIMA, a popular forecasting method, is effective for univariate time series data with trends and seasonality. This project focuses on implementing ARIMA to capture these temporal patterns and generate reliable forecasts.

**Steps for Time Series Sales Data Preprocessing:**

**Step 1: Load and Prepare the Data**

Import required libraries, load the dataset, convert the 'Month' column to datetime format, and set it as the index to structure it for time series analysis. This step ensures that the data is properly formatted for temporal operations and modeling.

import pandas as pd

import matplotlib.pyplot as plt

from statsmodels.tsa.arima.model import ARIMA

from pandas.plotting import autocorrelation\_plot

from statsmodels.tsa.stattools import adfuller

path = '/content/Autism.csv'

df = pd.read\_csv(path)

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

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

**Step 2: Check for Stationarity**

Use the Augmented Dickey-Fuller (ADF) test to statistically check if the time series is stationary. Stationarity is a crucial assumption for ARIMA modeling to work effectively.

result = adfuller(df['#Passengers'])

print(f'ADF Statistic: {result[0]}')

print(f'p-value: {result[1]}')

**Output:**

ADF Statistic: 0.8153688792060498

p-value: 0.991880243437641

**Step 3: Make the Series Stationary**

If the series is not stationary (p-value > 0.05), apply differencing.

df\_diff = df['#Passengers'].diff().dropna()

# Optional: plot autocorrelation to decide on ARIMA parameters

autocorrelation\_plot(df\_diff)

plt.title("Autocorrelation Plot")

plt.show()

A graph with blue lines

AI-generated content may be incorrect.

**Step 4: Build and Train the ARIMA Model**

Fit an ARIMA model to the time series using specified (p, d, q) values (e.g., (2, 1, 2)).  
The model learns from the past values and their relationships to forecast future points.

model = ARIMA(df['#Passengers'], order=(2, 1, 2))

model\_fit = model.fit()

print(model\_fit.summary())

**Step 5: Forecast and Visualize Results**

Forecast future values (e.g., next 12 months) and visualize them along with the original series. This helps to compare predicted values with historical data and assess the forecast's quality.

forecast = model\_fit.forecast(steps=12)

print("Forecasted values:")

print(forecast)

plt.figure(figsize=(10, 5))

plt.plot(df, label='Original')

plt.plot(forecast.index, forecast, label='Forecast', color='red')

plt.title('Superstores Passengers Forecast')

plt.legend()

plt.show()

**Output:**

Forecasted values:

1961-01-01 439.854513

1961-02-01 465.296054

1961-03-01 500.665761

1961-04-01 535.971659

1961-05-01 561.690109

1961-06-01 571.314405

1961-07-01 562.974263

1961-08-01 539.731146

1961-09-01 508.529538

1961-10-01 478.147818

1961-11-01 456.746809

1961-12-01 449.695642

Freq: MS, Name: predicted\_mean, dtype: float64

A graph showing the growth of passengers

AI-generated content may be incorrect.

**Result:**

The ARIMA(2,1,2) model effectively forecasted the next 12 months of superstores passenger traffic, showing a continuing upward trend in passenger numbers.