**Experiment 4**

**Name: SAP ID:**

|  |  |
| --- | --- |
| **Date:** |  |
| Aim | **Apply ARIMA - SARIMA models to predict / forecast the number of passengers travelling using aeroplanes based on time series data.** |
| Software | Colab |
| Pre-requisite ipa | Internet and required dataset |
| Theory | There are 2 ways to check for Stationarity of a Timeseries:  1. Rolling Statistics - Plot the moving avg(mean) or moving standard deviation to see if it varies with time. It is a visual technique.  2. ADCF Test - Augmented Dickey–Fuller test is used to gives us various values that can help in identifying stationarity. The Null hypothesis says that a Timeseries is non-stationary. It comprises of a Test Statistics & some critical values for some confidence levels. If the Test statistics is less than the critical values, we can reject the null hypothesis & say that the series is stationary. THE ADCF test also gives us a p-value. Acc to the null hypothesis, lower values of p is better.  What Is ADCF test?  In statistics and econometrics, an augmented Dickey–Fuller test (ADF) tests the null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity. It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models.  The augmented Dickey–Fuller (ADF) statistic, used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.  p value(0<=p<=1) should be as low as possible. Critical values at different confidence intervals should be close to the Test statistics value.  What is ARIMA model?  ARIMA(Auto Regressive Integrated Moving Average) is a combination of 2 models AR(Auto Regressive) & MA(Moving Average). It has 3 hyperparameters - P(auto regressive lags), d(order of differentiation) and Q(moving avg.) which respectively comes from the AR, I & MA components. The AR part is correlation between prev & current time periods. To smooth out the noise, the MA part is used. The I part binds together the AR & MA parts.  How to find value of P & Q for ARIMA ?  We need to take help of ACF(Auto Correlation Function) & PACF(Partial Auto Correlation Function) plots. ACF & PACF graphs are used to find value of P & Q for ARIMA. We need to check, for which value in x-axis, graph line drops to 0 in y-axis for 1st time. From PACF(at y=0), get P From ACF(at y=0), get Q  What is Exponential Smoothing?  Exponential smoothing is a rule of thumb technique for smoothing time series data using the exponential window function. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time. It is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality. Exponential smoothing is often used for analysis of time-series data. |
| Result |  |
| Conclusion | Write your detail understanding about the experiment. |