Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

Subject: Time Series Analysis

Experiment 6

(Autoregressive Integrated Moving Average)

Aim: Comparative Analysis of Autoregressive, Moving Average and Autoregressive Integrated Moving Average on a given dataset.

Theory:

Autoregressive integrated moving average—also called ARIMA(p,d,q)—is a forecasting equation that can make time series stationary with the help of differencing and log techniques when required. A time series that should be differentiated to be stationary is an integrated (d) (I) series. Lags of the stationary series are classified as autoregressive (p), which is designated in (AR) terms. Lags of the forecast errors are classified as moving averages (q), which are identified in (MA) terms.

A non-seasonal ARIMA model is called an ARIMA (p, d, q) model, where:
• p is the number of autoregressive terms.
• d is the number of non-seasonal differences needed for stationarity.
• q is the number of lagged forecast errors in the prediction equation.

Representation of p, d, q and Its Relevant Methods:

	p	d	q	Differencing	Method
ARIMA (0, 0, 0)	0	0	0	$y_t = Y_t$	White noise
ARIMA (0, 1, 0)	0	1	0	$y_t = Y_t - Y_{t-1}$	Random walk
ARIMA (0, 2, 0)	0	2	0	$y_t = Yt - 2Yt-1 + Yt-2$	Constant
ARIMA (1, 0, 0)	1	0	0	$\hat{Y}t = \mu + \phi_1 Y_{t-1} + \varepsilon$	AR(1): First-order regression model
ARIMA (2, 0, 0)	2	0	0	$\hat{Y}t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon$	AR(2): Second-order regression model
ARIMA (1, 1, 0)	1	1	0	$\hat{Y}t = \mu + Yt-1 + \phi 1\ (Yt-1 - Yt-2)$	Differenced first-order autoregressive model
ARIMA (0, 1, 1)	0	1	1	$\hat{Y}t = Yt-1 - \phi 1et-1$	Simple exponential smoothing
ARIMA (0, 0, 1)	0	0	1	$\hat{Y}t = \mu_0 + \varepsilon t - \omega_1 \varepsilon_{t-1}$	MA(1): First-order regression model
ARIMA (0, 0, 2)	0	0	2	$\hat{Y}t = \mu_0 + \varepsilon t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2}$	MA(1): Second-order regression model

	p	d	q	Differencing	Method
ARIMA (1, 0, 1)	1	0	1	$\hat{Y}t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon t - \omega_1 \varepsilon_{t-1}$	ARMA model
ARIMA (1, 1, 1)	1	1	1	$\Delta Y_t = \phi_1 Y_{t-1} + \varepsilon t - \omega_1 \varepsilon_{t-1}$	ARIMA model
ARIMA (1, 1, 2)	1	1	2	$\hat{Y}_t = Y_{t-1} + \phi_1 (Y_{t-1} - Y_{t-2}) - \theta_1 e_{t-1} - \theta_1 e_{t-1}$	Damped-trend linear Exponential smoothing
ARIMA (0, 2, 1) OR (0,2,2)	0	2	1	$\hat{Y}_t = 2\,Y_{t-1} - Y_{t-2} - \theta_1 e_{t-1} - \theta_2 e_{t-2}$	Linear exponential smoothing

ARIMA is a method among several used for forecasting univariate variables, which uses information obtained from the variable itself to predict its trend. The variables are regressed on its own past values. AR(p) is where p equals the order of autocorrelation (designates weighted moving average over past observations) z I (d), where d is the order of integration (differencing), which indicates linear trend or polynomial trend z. MA(q) is where q equals the order of moving averages (designates weighted moving average over past errors). ARIMA is made up of two models: AR and MA.

The Integration (I)

Time-series data is often nonstationary, and to make time-series stationary, the series needs to be differentiated. This process is known as the integration part (I), and the order of differencing is signified as d. Differencing eradicates signals with time, which contains trends and seasonality, so this series contains noise and an irregular component, which will be modelled only.

d(I) can be articulated algebraically:

Integral d Value	Formula(y_t)
d = 0	Y_t
d = 1	$Y_t - Y_{t-1}$
d = 2	$Y_t - 2Y_{t-1} + Y_{t-2}$

Lab Assignments to complete:

Perform the following tasks using the datasets mentioned. Download the datasets from the link given:

Link:
https://drive.google.com/drive/folders/1dbqJuZJULas76_Zzkqs-yRd2DbJReJup?usp=sharing

Colab Links: https://colab.research.google.com/drive/1-vwxWr31Bg6JpBna4MaTe7ABtxgcvvP3

Dataset 1: Facebook Stock Market Performance

1. Implement ARIMA (p, d, q) model on the given dataset.
    a. Plot a histogram and compare the values with N (0,1).
    b. Check for Stationarity.
    c. State the coefficients which has to be used for the appropriate model selected.
    d. Calculate the evaluation metrics (MSE, RMSE, MAPE and R2).
    e. Forecast the future values and plot Confidence Interval Upper bound and Confidence Interval Lower bound with respect to train, test and predicted.
    f. Analyse the actual data with predicted based on the plots:
        i. Standardize Residual
        ii. Histogram plus estimated density
        ii. Normal Q-Q
        iv. Correlogram

CODE:

https://colab.research.google.com/drive/11zyU4nyN9WHrTtxs7SGPmgveKJXG16H1?usp=sharing