

Stationary data: The data series is stationary, which means that the mean and variance should not vary with time.

A series can be made stationary by using log transformation or differencing the series(d).

Vector ARIMA

A Multivariate Time Series consist of more than one time-dependent variable and each variable depends only on its past values.

To deal with multi variate time series, one of the most popular methods is Vector Auto Regressive Integrated Moving Average models (VARIMA) that is a vector form of autoregressive integrated moving average (ARIMA) that can be used to examine the multivariate time series analysis[1].

Vector ARIMA model is commonly used machine learning model which is used to forecast the time series data. It is the merger of two components which are

- 1) **AR(Auto Regressive):** This model attempts to predict the future values based on past values. For this model, time series data should be stationary[2].

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t$$

Where:

y_t Is the value at time step t, c is a constant, ϕ_1 is a coefficient, and ϵ_t is a white noise error term with $\epsilon_t \sim N(0, \sigma^2)$.

From this model we will find the value p, which represents how many prior time steps to use in the time regression.

- 2) **MA(Moving Average):** It is the linear regression of the present value of the series against previous observed error terms [2].

$$y_t = c + \theta_1 \epsilon_{t-1}$$

Where:

y_t Is the value at time step t, c is a constant, ϕ_1 is a coefficient, and ϵ_t is a white noise error term with $\epsilon_t \sim N(0, \sigma^2)$.

From this model we will find the value q, which represents how many prior errors we need to consider.

Applications:

- 1) VARIMA model uses statistical analyses in combination with collected historical data points to predict future trends. so, data should be more reliable.
- 2) Consists of more components (such as AR, MA), so that the error will be less.

Limitations:

- 1) Data should be stationary for the model to work.
- 2) Long time span of data is needed for the model to forecast the future values effectively.

Steps to perform the algorithm:

- 1) **Loading the data:** Load the data, pre-process the data (divide the data into country wise) and split the data into train and test data.
- 2) **Make the series stationary:** Perform the Augmented Dickey–Fuller test and check the data is not stationary, do the difference operation(d) on the data until data becomes stationary.
- 3) **Determine p and q values:** We need to find p-value from Auto-correlation function (ACF) plot and q-value from Partial Auto-Correlation Function (PACF) plot.
- 4) **Fit the ARIMA Model:** Take the train data and p, q, d as parameter values and fit it into the model[4].
- 5) **Forecast values:** The model will forecast the future values by considering trained data and parameter values.
- 6) **Calculate MAPE:** Calculate mean absolute percentage error (MAPE) for each country from forecasted data and test data.

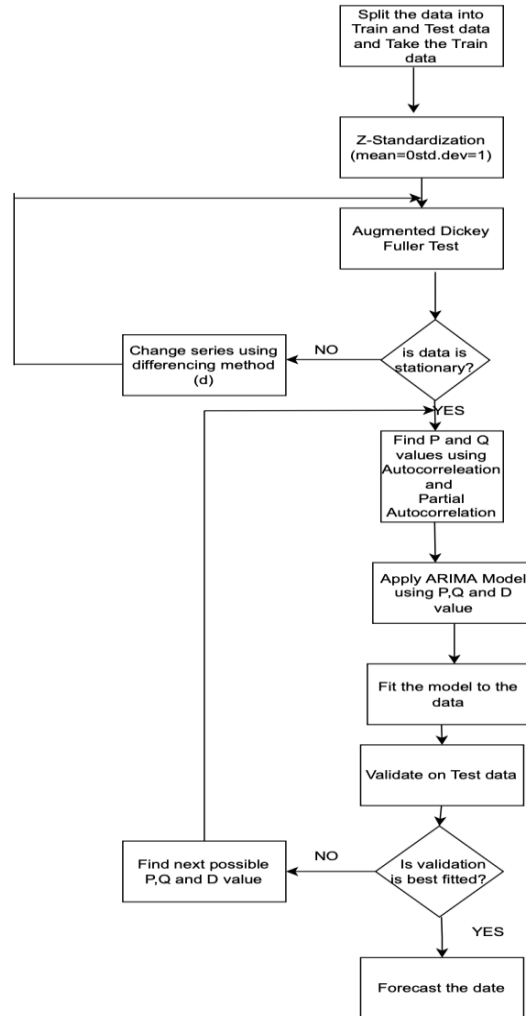


Figure: Flow chart of VARIMA model.

VARIMA Using Darts Library:

In the research we used Darts library to perform VARIMA. The main use is we no need to do (2) and (3) steps. It can directly forecast the data (5 step).

Results:

After calculating the MAPE for each country, we conclude that[3]

- 6 countries got less than 10% error.
- 10 countries got the error in between 10%-20%.

- 12 countries got the error in between 20%-30%.
- 12 countries got the error greater than 30%.

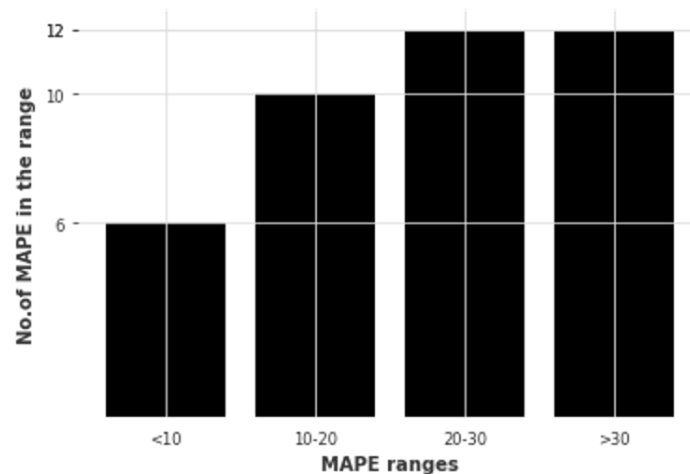


Figure: MAPE ranges of different countries.

Conclusion:

We are rejecting the algorithm because

1. Error is more compared to other algorithms.
2. The main drawback of algorithm is it is not able to find the hidden relations between the data which results to more error in forecasted data.
3. The historical data is very less which leads ineffective forecast of future values.

References:

1. Xin Chen, *A Multivariate Time Series Modeling and Forecasting Guide with Python Machine Learning Client for SAP HANA*, 06-May-2021, <https://blogs.sap.com/2021/05/06/a-multivariate-time-series-modeling-and-forecasting-guide-with-python-machine-learning-client-for-sap-hana/>
2. Lleyton Arton, *A Thorough Introduction to ARIMA Models*, 20 Jan 2021, <https://medium.com/analytics-vidhya/a-thorough-introduction-to-arima-models-987a24e9ff71>.
3. Darts, *Models of Varima*, https://unit8co.github.io/darts/generated_api/darts.models.forecasting.varima.html

4. Santanu Roy, Gouri Sankar Bhunia and Pravat Kumar Shit, *Spatial prediction of COVID-19 epidemic using ARIMA techniques in India*, 16 Jul 2020, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7363688/>