Forecasting COVID-19 Cases Using Time Series Models: A Study Across Countries and WHO Regions

The COVID-19 pandemic has impacted countries all over the world, but the number of cases, recoveries, and deaths has varied widely from place to place. Being able to understand and predict how case counts change over time is important for governments and health organizations so they can plan resources, respond effectively, and prepare for future waves.

Time series forecasting techniques, such as ARIMA and SARIMAX, are commonly used to model and predict trends in data that change over time. These models can help estimate how many cases might occur in the near future based on patterns observed in the past.

Objective

The goal of this project is to build and evaluate time series models to forecast the number of confirmed COVID-19 cases over time, at both the country and WHO region level. The focus will be on using ARIMA and SARIMAX models to capture trends and seasonal patterns in the data and to compare their performance.

Key Questions

- Can time series models reliably predict daily or weekly confirmed COVID-19 cases for a given country?

- How do forecasts differ across WHO regions?

- Which model, ARIMA or SARIMAX, works better for this type of data?

Scope

- Explore the data to identify trends, seasonal effects, and other patterns.

- Build time series models to forecast future confirmed case counts.

- Evaluate the models using appropriate metrics, such as RMSE and MAPE.

- Compare the results and discuss which model performs best and why.

- Visualize the forecasts and explain what they could mean for public health planning.

Deliverables

- A cleaned and prepared dataset ready for modeling.

- Visualizations showing trends, patterns, and model forecasts.

- Working time series models (ARIMA and SARIMAX) with performance evaluations.

- A written report or presentation summarizing the approach, findings, and recommendations.

Dataset
Link: https://drive.google.com/file/d/1TJdxCYLWOIDB4UPzXiLIOAfe1Sibmi6r/view?usp=sharing