**Quantitative model development for time series forecasting air quality and pollen**

Task 2 - Forecasting Air Quality Model:

SARIMA Assumption: Let’s assume that we have to build a time series forecasting model on an hourly dataset of particulate matter, PM.

ARIMA is one of the most widely used time series models and it stands for Autoregressive Integrated Moving Average. ARIMA models are denoted with the notation ARIMA(p, d, q). These three parameters account for seasonality, trend, and noise in data.

However, ARIMA does not support time series with a seasonal component. Time series data such as the air quality does contain seasonal components and to model such data we can use an extension to ARIMA that supports the direct modelling of the seasonal component of the series. It is called SARIMA.

A SARIMA model requires selecting hyperparameters for both the trend and seasonal elements of the time series.

Trend Elements

They are the same as the ARIMA model:

● p: Trend autoregression order

● d: Trend difference order

● q: Trend moving average order

Seasonal Elements

There are four seasonal elements:

● P: Seasonal autoregressive order

● D: Seasonal difference order

● Q: Seasonal moving average order

● m: The number of time steps for a single seasonal period

The SARIMA time series forecasting method is supported in Python via the Statsmodels library.

Q. A What would be the key challenges for building a forecasting model for air quality?

The key challenges for building a forecasting model that one might encounter are listed below:

● Inaccurate readings from the sensors

● Climate change, can’t rely completely on historical data

● Model overfitting

● Noisy data Q.

B How would the difference between output 1 or 2 impact the complexity when building the model?

For output 1 we would have to build a single model for a specific location. However, for output 2 we would have to build multiple models because there would be multiple locations on a map and since we have to forecast for multiple time periods, it further increases the model complexity.

Hence, building a model for output 2 would be more difficult as compared to the case of output 1.

Task 3 - Forecasting Pollen

QA. How would building a model for pollen forecasting differ from building a model for forecasting air quality?

Pollen forecasting is quite different from forecasting air quality on so many fronts. Most importantly it takes into account multiple environmental factors that might affect the timing of pollen release by different plants. Some key factors that help in pollen forecasting are listed below:

● Temperature

● Precipitation

● Humidity

● Wind Direction

● Wind Speed

In terms of machine learning, we call it a multi-variate time series data.

To model such multi-variate time series, we can use regression models to deep learning models such as Recurrent Neural Networks where there are multiple input variables.