# **INTERPRETATIONS:**

**ARIMA:**

Before reading the interpretations, we would ask you to read through the basic understanding of the model in Time Series Machine Learning Models. If you have gone through it, you would know that ARIMA has three hyperparameters – p, q, d and it is important to have the perfect combination of these three hyperparameters so that ARIMA works optimally on the data. In weather data, p and q is assigned the value you choose as window size. On the other hand, d is computed by counting the number of times differencing is done so that the correlation score is reduced below 0.05, i.e., the stationarity is removed completely. This is done by the adfuller function of statsmodel. You can refer to the code snippet of ARIMA to see how these are implemented.

**Results observation**

If you look at the results, you will notice that ARIMA doesn’t perform as good as the other models. Temperature and humidity have seasonality and ARIMA doesn’t perform well with seasonal data. On the other hand, precipitation has an interesting result in comparison to temperature and humidity. We observe that precipitation is not seasonal and is quite random. What makes this interesting is that the models don’t perform well either like ARIMA except SVR.

**Results analysis**

A possible reason for this weak performance is seasonality of the trend of temperature and thus the choice of p and q, which in this case would be optimal if chosen as 365, i.e., if ARIMA can peek into the past one year’s information, it would be easier for it to make a proper decision. However, the increase in exponential time complexity of the model with higher p and q values led us to choose smaller values for window which directly reflects in the prediction and the mean absolute score. Another important aspect is that ARIMA heavily depends on the past information and due to the data’s highly changing nature, it is important that there is a large amount of data. In this case, we don’t have enough data

In conclusion, in recent years, other machine learning techniques (e.g., Linear Regression, Random Forest, Support vector machines) and other deep learning techniques (e.g., RNN and LSTM) have replaced these classical machine learning models due to their reduced complexity and reduced need for larger dataset for similar performance. This can be seen in the mean absolute square error comparison with other models especially for humidity and temperature.

**SARIMA:**

Before reading the interpretations, we would ask you to read through the basic understanding of the model in Time Series Machine Learning Models. If you have gone through it, you would know that SARIMA has three non-seasonal hyperparameters – p, q, d and four seasonal hyperparameters – P, Q, D, m.

For our experiments, we just tested the seasonal hyperparameters i.e., leaving the non-seasonal parameters as zeros. It is important to have the perfect combination of these hyperparameters so that SARIMA works optimally on the data. In weather data, P, Q, and d are assigned the value of 1, and m defines the seasonal cycle where we want the model to look. In this case, m is assigned the value you choose as window size. For our data, the ideal value of m would be 365, so the model looks back a year to get an idea of what the weather was at this time of the year previous year. But due to SARIMA’s increased complexity, we stick to smaller m values.

**Results observation**

If you look at the results, you will notice that SARIMA doesn’t perform as good as the other models. Although SARIMA performs well with seasonal data, in this it doesn’t perform well with temperature and humidity. On the other hand, precipitation has an interesting result in comparison to temperature and humidity. We observe that precipitation is not seasonal and is quite random. What makes this interesting is that the models don’t perform well either like SARIMA except SVR.

**Results analysis**

A possible reason for this is the choice of m, which in this case would be optimal if chosen as 365, i.e., if SARIMA can peek into the past one year’s information, it would be easier for it to make a proper decision. However, the increase in exponential time complexity of the model with higher m values led us to choose smaller values for window which directly reflects in the prediction and the mean absolute score. Another important aspect is that SARIMA heavily depends on the past information and due to the data’s highly changing nature, it is important that there is a large amount of data. In this case, we don’t have enough data.

In conclusion, in recent years, other machine learning techniques (e.g., Linear Regression, Random Forest, Support vector machines) and other deep learning techniques (e.g., RNN and LSTM) have replaced these classical machine learning models due to their reduced complexity and reduced need for larger dataset for similar performance. This can be seen in the mean absolute square error comparison with other models especially for humidity and temperature.

**RNN**

*RNN LSTM, better performance with more features*