Data prediction using ARIMA Model

**Objective**

To design a model for forecasting or predicting future data based on collected data.

**Problem statement**

We live in a world where environmental monitoring is essential in our daily lives. Weather monitoring is extremely important and is used in a variety of fields, including agriculture and industry. Weather monitoring allows us to keep an eye on factors like temperature, humidity, atmospheric pressure, light rain, wind speed, and wind direction.

Having knowledge of possible weather parameters is and advantage in the field of agriculture, industrial and the likes.

**Tools used**

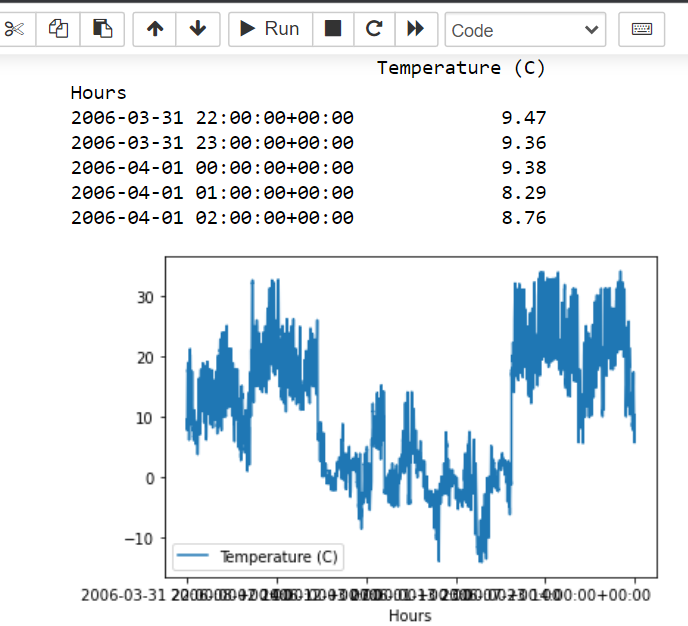
* Microsoft excel
* Jupyter notebook

**Methods used**

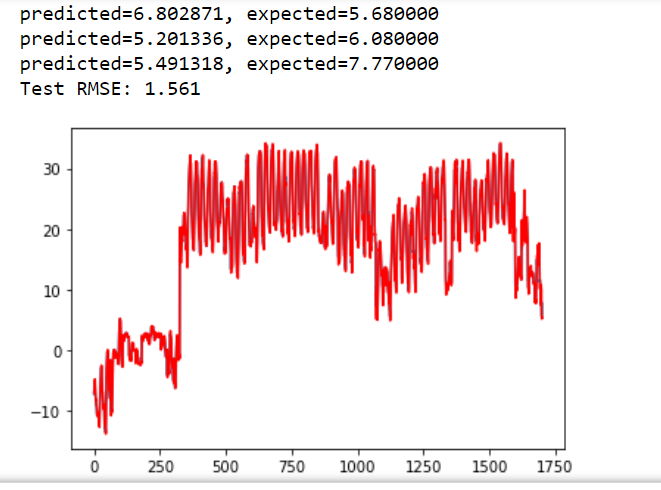
* Downloaded weather data from [Historical Hourly Weather Data 2012-2017 | Kaggle](https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data) and saved locally as csv file
* Used python in data cleaning proceedure (ie removed unwanted colums, dropped empty rows, removed duplicates and standardized outliners)
* Read the data in excel using only the temperature and hour parameters and I indexed the hour column
* I plot the temperature to time graph using the first half of the dataset
* Using sqlearn library and the arima model I developed a training algorythm using the initial first half of the data set to train and predict the second half
* The model responded with numerous running with a minimal rms.

**Results**

The figures below are the graphical representation of data demonstrating the prediction and the model’s accuracy after feeding it a series of example data.



**Fig 1 Original temperature data visualization**



**Fig 2 Predicted tempearature visualization**

Fig.1 and Fig.2 are a full dataset of a weather data (temperature values) collected and fed to the model. Fig.1 is the initial half of dataset and Fig.2 is the final half displayed together with the predicted values of the data. After a series of training the model, this example had the smallest root mean square, therefore, making the predicted data in Fig.2 very close to the future values of Fig.1 which appear in blue in Fig.2.

**Conclusion**

The model is designed to accept data and predict the future value of that data with the closest precision as possible.

**Recommendation**

Upon completion of the project, I recommend that a webserver is built for collection, storage and monitoring of these data values from an IoT device.

And I recommend that the model is linked to this webserver that will be able to automatically forecast future values collected by the IoT device based on a machine learning algorithm.