

## IoT class project [Fall 2021]

**Project report deadline: Nov 14, 2021 , 8 PM IST (Late submissions not accepted)**

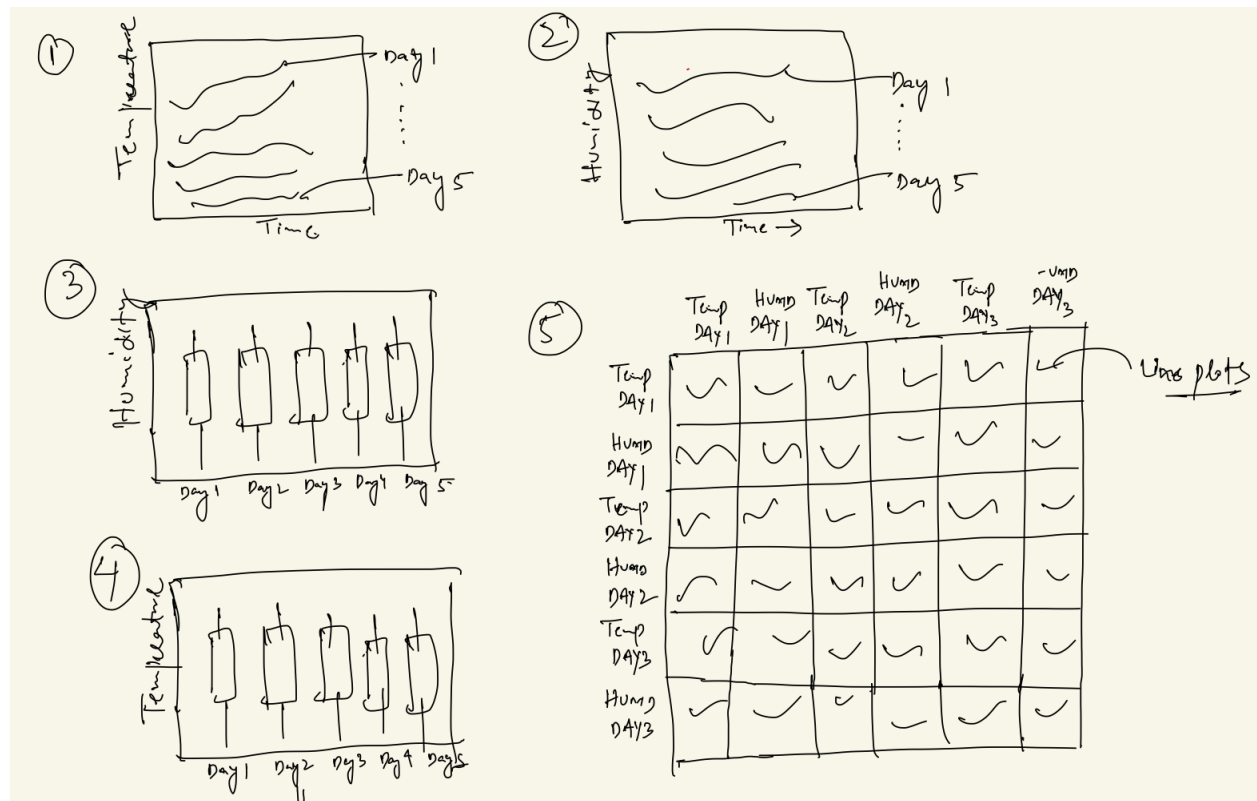
**Project evaluation: Nov 15 week**

**Don't copy/share your work. Such submissions will get a penalty of -20**

The project comprises of multiple components, i.e., plotting and analysis. The following sections provide the details of each component.

### Component # 1: Plotting

**Task:** Use Weather\_data\_2sites dataset (found on Canvas) and plot the following five plots with your favourite programming language. I suggest to use Python pandas.



Use timestamp, temperature\_site1, and humidity\_site1 columns present in the dataset for the plotting. Ignore the remaining columns for the plotting purpose. The dataset contains data of approximately three months and you need five days of data for this component. Use the following strategy to select the five days - Use the last two digits of your roll number to select the five days of the dataset. For example, a student with a roll number 18199 will plot data of 5 days starting from day 19 (= 9+9+1) onwards of any month. Similarly, a student with a roll number 18100 will plot data of 5 days starting from day 1 (=0+0+1) onwards of any month .

**Plot 1 and 2:** Line plots for temperature and humidity data showing 5 lines in each plot corresponding to 5 days of data. Time on the X-axis should show day hour. Plot legend must show the day number (i.e., month day) used for the plotting. For example, if you used days 15, 16, 17, 18, and 19 for the plotting then the plot should display this; **Plot 3 and 4:** Box plots on temperature and humidity data such that each box corresponds to one day's data. X axis labels should reflect the days used for the plotting; **Plot 5:** Line grid plot for any 3 days of data where labels on the plot are temperature day[x], humidity day[y]. Axis labels must show which days data you have used for plotting.

#### **Evaluation:**

1. The plots must be visible and labeled properly (axis labels and legends) [ 10 points, two for each plot]
2. Show programming script used for each plot separately [10 points, two for each plot]
3. Explain each plot in 2-3 sentences, i.e, what does the plot convey [10 points, two for each plot]

## **Component # 2: Analysis**

Use Weather\_data\_2sites dataset to create a new dataset, Syn\_weather, consisting of 5 variables, i.e, temperature, humidity, day-minutes, day-of-the-week, previous-temperature. Such that the dimensions of the Syn\_weather will be 5779 x 5. Use the Sys\_weather dataset and perform the following tasks

1. Create a regression model, M1, predicting temperature using variables humidity, day-minutes, day-of-the-week, previous-temperature. Use  $\frac{2}{3}$  of data without replacement for training and rest for  $\frac{1}{3}$  for testing. Report RMSE. [10 points]

2. Create a regression model, M2, predicting temperature using variables humidity, day-minutes, day-of-the-week, previous-temperature. Use 10 fold Cross-validation approach and report RMSE. [10 points]
3. Use step-wise regression to find the top-2 predictors in M1. Explain briefly your findings. [5 points]
4. Compute correlation coefficient of all variables in Syn\_weather. Describe your findings. [5 points]
5. Plot actual vs predicted values for both M1 and M2 side-by-side. The plot should be properly labeled. Describe your observations. [5 points]

Hint: Use pandas to extract minutes, week day from the timestamps provided in the original dataset.

**Note:** Make sure to submit code, plots, results, findings in the report. Well-formatted reports will get **5** bonus points.