**EE100 – Engineering Laboratory Spring 2022**

**Finding the Blue Skies: Analyzing Lahore’s Air Quality Data**

**A picture containing outdoor, sky, sunset, sun

Description automatically generated**

**Instructor:**

Dr. Momin Uppal

**Group Members:**

Aiman Mubasher 2024-10-0312

Waleed Arshad 2024-10-0303

**Overview:**

Over the years, the city of Lahore has continued to make records among the worst Air quality recorded for a city. Especially during winter, Lahore remains at the top for the Worst Air Quality in AQI. In this project, we plan to analyze the data we have on some sensors installed at specific locations all over the city. In our analysis and report, we will be using data from thirteen sensors installed in the city by both government and crowd-sourced sensors, while the USEmbassy sensor will be used as our focal data.

Table 1 shows the details of the detectors for which we analyzed the data.

**Table

Description automatically generated**

**Task 1: Displaying the location of sensors**

In this task, we have labeled the locations of the sensors using the values of longitudes and latitudes provided to us in the file (Locations.csv). Here, we used the MATLAB built-in tool *readtable* to read values from the csv file. We used the *geoscatter* plot tool to pinpoint the locations of the stations.

A picture containing graphical user interface

Description automatically generated

**KEY:**

Blue – USEmbassy (AirNow)

Yellow – Cheap Crowdsourced sensors(PurpleAir)

Red – (EPD)

**Task 2: Calculating Distances**

In this task, we calculated distances from the US Embassy sensor with respect to the other twelve sensors. Since there are 13 sensors in total, we have calculated 78 unique distances. The haversine formula has been used to compute the distance between two points whose lat/long are given. The observations are plotted in the following bar chart.

**Chart, bar chart

Description automatically generated**

**Task 3: Computing Correlation Coefficients**

The process of establishing a relationship or connection between two or more things is known as its correlation (Oxford). We found correlations between USEmbassy data with all other sensors in this task. This correlation determines the data integrity of different sensors. The data was then plotted into a bar graph in this task, as shown below. USEmbassy station has been used as a reference-grade.

**Chart, bar chart

Description automatically generated**

**Task 4: Calibrating the Sensor Readings**

In this task, we do linear calibration on the data of all the sensors except EPD using the USEmbassy data as our calibration point. Implementation is done on MATLAB.

We used the Polyfit function for estimation, which returns the coefficients for a polynomial p(x) of degree n that is the best fit (in a least-squares sense) for the data in y. We have used degree 1 of the polynomial function, a linear graph. We did not opt for a higher degree polynomial to avoid overfitting data.

For the shortcoming, we believe the estimation might not be accurate, as, for most of the dates, data was not available, thus giving us an inaccurate estimate. Having a smaller data set could result in less accuracy in our prediction.

**Task 5: Estimating the missing values at LUMS**

We implemented inverse distance weighting using the formula provided to us in the manual. We used that formula to estimate the missing values. After running some Matlab commands, the values where no data is available are filtered out. Here, we observe an upward trend in the AirQuality Index.

To determine the accuracy of our estimation, one method that can be used is to divide the data into test and train data. We can use the same algorithm we designed to estimate values from December 2021 onwards. Then we can calculate the mean squared error using a mathematical formula and finally find the accuracy. This accuracy value will help us determine how accurate our estimate is. If our value is greater than 90%, our algorithm is reliable. We can also manually compare our estimated values with the data from December 2021 onwards and see how close our values are to the original values.

**Chart, histogram

Description automatically generated**

**Work Division:**

We have equally divided work among ourselves. We have worked together on all the tasks by sitting together and working on the same system. Sometimes, one member will code, and the other will manually work on the pseudo-code, and then the other way around.