Dyson School of Design Engineering | MEng Design Engineering

**Module Exam**

|  |  |
| --- | --- |
| Module code and Name | DE4-SIOT Sensing & IoT |
| Student CID |  |
|  |  |
| Assessment date | **10th Jan 2019** |
|  |  |

**Presentation URL (publicly accessible link):**

**Code & Data (publicly accessible link):**

# Coursework 1: Sensing

### Introduction and objectives

According to an article (Ogden, 2016), “In London alone, bad air quality is thought to kill nearly 10,000 people a year.” Currently, it is estimated that air pollution causes 15% of Chronic Obstructive Pulmonary Disorder which is predicted to the third leading cause of premature death worldwide by the World Health Organisation. A King’s College London study estimates that London’s two main pollutants (nitrogen dioxide, NO, and fine particulates) are responsible for 5, 900 premature deaths a year (Excell, 2015).

In this part of the project, the relationship between the how much people care about air pollution and the air quality throughout time is explored. The number of tweets about air pollution over a period is used to represent the awareness and care of the people. A periodic measure of the different pollutants in the air is used to represent the changes in air quality throughout time.

### Data Collection

The data sources used are:

(i) OpenAQ API, which provides open air quality data (OpenAQ, 2018). This is used to collect real-time pollutants concentrations of different locations.  
(ii) Twitter Developer’s Standard Search API, which searches through tweets in Twitter (Twitter, 2018), a popular social media platform. This is used to collect information of real-time twitter posts with the key phrase “air pollution”.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Timestamp** | **Location** | **pm25** | **pm10** | **no2** | **o3** | **co** |
| 2019-01-04T14:00:00.000Z | Causeway Bay | 46.7 | 64.3 | 165.8 | 2.7 | 852.8 |
| 2019-01-04T14:00:00.000Z | New Territories | 53.7 | 95.2 | 102.5 | 2 | 877.6 |
| *Sample of the data collected from OpenAQ API* | | | | | | |

Air quality being location dependent, different locations were selected as focus for measuring the air quality. The locations were selected based on the places that have higher tweeting frequencies (more tweets over a period) so that the changes and trends in tweeting frequency can be more obvious, and places that may yield more interesting data about pollution. Data collection occurred throughout ten days (day and night) due to the time dependency of tweet rates, since there are various time differences in the selected locations.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Timestamp** | **Text** | **User** | **Tag** | **Location** | **Specific Location** | **General Location** |
| Fri Jan 04 14:45:13 | RT @BkPhilanthropy: Mercury pollution in the air we breathe m is down 81% because of regulations. https://t.co/uhXmCAW5dm | KCSunshineMom |  | Geeks Resist HQ | other | - |
| Fri Jan 04 14:45:02 | Air is OK near Croydon - Park Lane (Pollution Low : 1) | breathinglondon |  | London | London | UK |
| *Sample of the data collected from Twitter API* | | | | | | |

Ideally, the location where each tweet is posted is recorded as the associated location. However, the Twitter API can only provide that information for geo-tagged posts, which is the minority of posts. Instead, the location that is set on the tweeter’s profile, which is provided for each tweet, is used. Further programming was used to categorise the tweets to the chosen different general locations, which is used to relate with the locations used for air quality data collection.

To prevent aliasing, the quantities of different air pollutants data are sampled at the maximum sampling frequency rate, which is a sampling period of one hour. OpenAQ API only provides new measured values every hour. The tracking of tweets about air pollution is recorded in real-time with a time stamp. This means that the fixed period chosen to define the tweet rate (number of tweets across a fixed period) can be decided after the data collection.

The data were stored using the Google Sheets API to automatically store the (near-) real-time streaming information into a Google Spreadsheet, which is stored in the cloud.

### Time-series Data Analysis

For initial data processing, the different pollutants data for the individual locations were grouped to give average data for their respective countries/large city. For further simplicity, the measured pollutant concentrations were converted into an Air Quality Index (AQI) for each four general locations. Inspired by how the U.S EPA AQI was calculated: the largest IAQI (Individual AQI), which is the index for each pollutant, was chosen on an hourly basis as the current AQI. IAQI was calculated as such . Missing data were filled with the most recent previous data. Large time gaps in the data collection were removed to not affect the regression.

[Graphs with ACF (AutoCorrelation Function)]

<https://www.epa.vic.gov.au/your-environment/air/air-pollution/air-quality-index/calculating-a-station-air-quality-index>

* Is there a trend, meaning that, on average, the measurements tend to increase (or decrease) over time?
* Is there seasonality, meaning that there is a regularly repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, and so on?
* Are their outliers? In regression, outliers are far away from your line. With time series data, your outliers are far away from your other data.
* Is there a long-run cycle or period unrelated to seasonality factors?
* Is there constant variance over time, or is the variance non-constant?
* Are there any abrupt changes to either the level of the series or the variance?

### Discussion

* Assumption with location
* Time gaps and data quantity

### Conclusion

## Coursework 2: Internet of Things

### Data analytics, inferences and insights

Although there are problems with air pollution, some people are unaware of the impact. A London study (Taylor and Laville, 2017) showed that only one in ten British adults rated air quality as poor despite the country exceeding the legal limits of air pollution. An article that describes air pollution as a “silent killer” emphasizes the unfamiliarity of Air Quality Index (AQI) to some people in Vietnam (Dat, T., 2018). The effects of air pollution are difficult to comprehend due to it being invisible.

### Data Interaction/visualisation/actuation platform

Gamification, social network (see tweets)

<https://www.fastcompany.com/3049540/5-ways-to-convince-people-to-actually-do-something-about-climate-change>

visualising

<http://www.carbonvisuals.com/blog/2016/2/3/a-breath-of-fresh-air-visualising-air-pollution>

<https://www.fastcompany.com/3062129/these-pollution-sensitive-shirts-visualize-the-filthy-air-youre-breathing>

### Discussions on the important aspects of the project

### Avenues for future work and potential impact

## References

Ogden, L.E., 2016. The tiny changes air pollution makes inside you. *BBC*. 11 February. [Viewed 14 November 2018]. Available from: <http://www.bbc.com/future/story/20160210-the-tiny-changes-air-pollution-makes-inside-you>.

Excell, J., 2015. The lethal effects of London fog. *BBC.* 22 December. [Viewed 14 November 2018]. Available from: <http://www.bbc.com/future/story/20151221-the-lethal-effects-of-london-fog>

Taylor, M., and Laville, S., 2017. British people unaware pollution levels air they breathe -study. [Viewed 14 November 2018]. Available from: <https://www.theguardian.com/environment/2017/mar/01/british-people-unaware-pollution-levels-air-breathe-friends-earth>

Dat, T., 2018. Air pollution is Vietnam’s silent killer. *Vietname Investment Review.* 26 February. [Viewed 26 December 2018] Available from: <https://www.vir.com.vn/air-pollution-is-vietnams-silent-killer-56542.html>

OpenAQ, 2018. *Open AQ Platform API.* [online] Available at: <https://docs.openaq.org/> [Accessed 23 December 2018].

Twitter, Inc., 2018. *Standard Search API.* [online] Available at: <https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html> [Accessed 14 November 2018].