**Minutes  
Machine Learning Model Development  
Meeting No. 9**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Location** | Swinburne Hawthorn Campus  Room EN4001 | | | | |
| **Date & Time** | Tuesday, September 17, 2024 8:30am | | | | |
|  |  | | | | |
| **Contributors** | Henry Richardson Matthew CrossSeth Kalantzis | | | *Meeting Facilitator Secretary* | |
|  | | | | | |
| **Action Items from Previous Meeting** | | | | | |
| **Item** | | **Discussion Points** | | | |
| **Attempt cleaning of data individually**  ***Owner:*** *All*  ***Status:*** *In-Progress* | | * Identified the following problem points over the weekend:   + Air Quality Dataset contains OZONE hourly and OZONE 4-hour average readings. Need to select only one when exporting data from source.   + Many locations begin or stop collecting data at certain dates. Need to consider threshold for including certain locations in the analysis.   + NSW Health Stats has the ability to break data down by Local Health District. Team had completed initial cleaning and analysis over the weekend using aggregated, state-wide data. Need to go back and consider local areas instead, as this may line up to how the air quality data is recorded. | | | |
|  | | | | | |
| **New Items** | | | | | |
| **Item** | | **Discussion Points** | | | |
| **Limiting Scope**  ***Owner:*** *Matthew*  ***Status:*** *Completed* | | * NSW Health Stats offers a significant range of data. Decision was made to limit the scope of analysis to how air quality impacts respiratory health. This still provides the team with about 10 different datasets to analyse. | | | |
| **Additional datasets / areas for analysis**  ***Owner:*** *Seth*  ***Status:*** *In-Progress* | | * Potential to add additional datasets with local emissions by factories if that data is available. Need to research. * Will return to this idea after initial analysis of existing datasets are completed. | | | |
| **Need for consistent datasets within the team for analysis**  ***Owner:*** *Matthew*  ***Status:*** *In-Progress* | | * The breadth of variables available when downloading datasets from NSW Health Stats and Air Quality NSW mean there is the potential for the team to download slightly different datasets if doing so individually. * Henry and Matthew to work on downloading datasets and uploading them to the team repository. * Information regarding parameters selected, website source, and date, to be included with the uploaded datasets. | | | |
|  | | | | | |
| **Action Items** | | | | | |
| **Action Item** | | | **Owner** | | **Deadline** |
| * **Add datasets to the team’s GitHub repository** | | | *Henry, Matt* | | *September 17, 2024* |
| * **Clean data and complete pre-processing** | | | *All* | | *September 20, 2024* |
|  | | | | | |
| **Next Meeting** | | | | | |
| **Location** | Swinburne Hawthorn Campus  Late Lab | | | | | |
| **Date & Time** | Friday, September 20, 2024 4:00pm | | | | | |



|  |  |  |
| --- | --- | --- |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Henry Richardson  *Meeting Facilitator* | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Matthew Cross  *Secretary* | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Seth Kalantzis  *Attendee* |