**Minutes   
Machine Learning Model Development  
Meeting No. 11**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Location** | Swinburne Hawthorn Campus Room EN401 | | | | |
| **Date & Time** | Tuesday, September 24, 2024 8:30am | | | | |
|  |  | | | | |
| **Contributors** | Henry Richardson Matthew CrossSeth Kalantzis | | | *Meeting Facilitator Secretary* | |
|  | | | | | |
| **Action Items from Previous Meeting** | | | | | |
| **Item** | | **Discussion Points** | | | |
| **Continue to pre-process data.**  ***Owner:*** *Matthew*  ***Status:*** *In-Progress* | | * Decision to normalization and splitting of data to be conducted independently with each individual model, to best fit the needs of the particular analysis. * All data cleaned, processed and integrated. * Additional work to standardise formatting of notebooks and comments. | | | |
| **Complete the regression model**  ***Owner:*** *Henry*  ***Status:*** *In-Progress* | | * Strong progress on linear regression model. * Identified that out initial hope of finding very strong correlations may have been too optimistic. | | | |
| **Complete the classification model**  ***Owner:*** *Seth*  ***Status:*** *In-Progress* | | * Classification proved challenging for the data collected for this assignment. * Focus was on linear regression and time series analysis during the week. * Will continue to work on classification model, specifically KNN. | | | |
| **Complete the clustering model**  ***Owner:*** *Matthew*  ***Status:*** *In-Progress* | | * Initial attempts made on K-means, DBSCAN and hierarchical clustering analysis. * More work to be done to compare metrics and determine most appropriate model. | | | |
|  | | | | | |
| **New Items** | | | | | |
| **Item** | | **Discussion Points** | | | |
| **File Paths Too Long**  ***Owner:*** *Henry*  ***Status:*** *In-Progress* | | * Identified issues with pulling files from git as the file paths had become too long. * Seemed to only be an issue for Henry’s machine. Matthew and Seth able to pull data successfully. * Henry downloading and uploading files manually as a temporary solution. | | | |
| **Threshold for Strong Correlation**  ***Owner:*** *Henry*  ***Status:*** *Complete* | | * Initially, the team had hoped to identify a strong correlation between certain pollutants and health outcomes. Initial expectations were to find r2 scores of at least 0.7 – 0.8. * The teams analysis, and subsequent research into the field, confirmed that this was unrealistic. Expectations of correlation have been revised. | | | |
|  | | | | | |
| **Action Items** | | | | | |
| **Action Item** | | | **Owner** | | **Deadline** |
| * **Update comments in data cleaning files** | | | *Matthew* | | *Sept 27, 2024* |
| * **Complete the regression model** | | | *Henry* | | *Sept 27, 2024* |
| * **Complete the classification model** | | | *Seth* | | *Sept 27, 2024* |
| * **Complete the clustering model** | | | *Matthew* | | *Sept 27, 2024* |
|  | | | | | |
| **Next Meeting** | | | | | |
| **Location** | Swinburne Hawthorn Campus Library, Level 3 | | | | | |
| **Date & Time** | Friday, September 27, 2024 2:00pm | | | | | |



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