

# Minutes

## Machine Learning Model Development Phase

Meeting No. 6

Location	Swinburne Hawthorn Campus Room EN401
Date & Time	Tuesday, August 27, 2024 8:30am

Contributors	Henry Richardson Matthew Cross Seth Kalantzis	Meeting Facilitator Secretary
--------------	---	----------------------------------

### Action Items from Previous Meeting

Item	Discussion Points
All outstanding items related to Milestone One.  Owner: Matthew Status: Completed	<ul style="list-style-type: none"><li>Project report, meeting minutes, and team contribution form all submitted on Sunday, August 25 by Matthew.</li><li>Project phase one officially complete.</li></ul>

### New Items

Item	Discussion Points
Review of Phase One  Owner: Matthew Status: Completed	<ul style="list-style-type: none"><li>Discussion around contributions of each team member. Each team member indicated that they were happy with the distribution of tasks and work completed by other team members.</li><li>Discussion around quality of work. Each team member indicated that they were happy with the quality of work produced by each team member.</li><li>Discussion around communication. Each team member indicated that they were happy with the quality and frequency of communication amongst the team in Project Phase One.</li><li>Overall, team very satisfied with the collaboration efforts from phase one. No major areas of improvement identified for phase two.</li></ul>
Identification of preliminary work required for Phase Two  Owner: Henry Status: In-Progress	<ul style="list-style-type: none"><li>The team agreed upon the following tasks to be completed over the coming week:<ul style="list-style-type: none"><li>Collate a number of potential data sources for review.</li><li>Brainstorm additional feature sets/ problem statements for the project to solve. (The current problem statement of predicting property prices identified as too simplistic to achieve high marks)</li></ul></li></ul>

Action Items

Action Item	Owner	Deadline
• Create a list of potential data sources.	All	September 03, 2024
• Create a list of additional features/ problems to solve.	All	September 03, 2024

Next Meeting

Location	Swinburne Hawthorn Campus Late Lab
Date & Time	Tuesday, September 03, 2024 1:00pm

*Henry Richardson*  
Henry Richardson  
*Meeting Facilitator*

*Matthew Cross*  
Matthew Cross  
*Secretary*

*Seth Kalantzis*  
Seth Kalantzis  
*Attendee*

# Minutes

## Machine Learning Model Development Phase

Meeting No. 7

Location	Swinburne Hawthorn Campus Room EN401
Date & Time	Tuesday, September 03, 2024 8:30am

Contributors	Henry Richardson Matthew Cross Seth Kalantzis	Meeting Facilitator Secretary
--------------	---	----------------------------------

### Action Items from Previous Meeting

Item	Discussion Points
Create a list of potential data sources.  <i>Owner: All</i> <i>Status: Overdue</i>	<ul style="list-style-type: none"><li>Each team member had completed a brief search for potential datasets, and collectively had found only one viable dataset (Melb Housing Prices, Kaggle).</li><li>Identified the need for multiple datasets to extend analysis and meet advanced marking criteria.</li><li>Each team member to research additional datasets for next meeting.</li></ul>
Create a list of additional features/ problems to solve.  <i>Owner: All</i> <i>Status: Paused</i>	<ul style="list-style-type: none"><li>Features to be derived from the identified datasets. Task delayed until additional datasets are identified.</li></ul>

### New Items

Item	Discussion Points
Prioritization of other tasks.  <i>Owner: All</i> <i>Deadline: n/a</i>	<ul style="list-style-type: none"><li>The team identified that prioritizing tasks from other subjects now could allow for more focused attention on this project closer to the deadline once appropriate datasets had been identified. No immediate tasks were set for the semester break period.</li></ul>

### Action Items

Action Item	Owner	Deadline
<ul style="list-style-type: none"><li>Find three potential datasets for the team to review.</li></ul>	All	Sept 13, 2024

Next Meeting

Location	Swinburne Hawthorn Campus Late Lab
Date & Time	Friday, September 13, 2024 4:00pm

*Henry Richardson*  
Henry Richardson  
*Meeting Facilitator*

*Matthew Cross*  
Matthew Cross  
*Secretary*

*Seth Kalantzis*  
Seth Kalantzis  
*Attendee*

# Minutes

## Machine Learning Model Development Phase

Meeting No. 8

Location	Swinburne Hawthorn Campus Late Lab
Date & Time	Friday, September 13, 2024 4:00pm

Contributors	Henry Richardson Matthew Cross Seth Kalantzis	Meeting Facilitator Secretary
--------------	---	----------------------------------

### Action Items from Previous Meeting

Item	Discussion Points
<b>Find three potential datasets for the team to review.</b>  <i>Owner: All</i> <i>Status: Completed</i>	<ul style="list-style-type: none"><li>Team completed a thorough search for appropriate housing data.</li><li>Matthew created spreadsheet with features in all datasets he explored. The only viable dataset only contained information from 2016-2018.</li><li>Henry explored datasets outside of Australia. Identified a few American housing datasets which had potential. Team exploration of these datasets found that they either didn't contain specific features of the houses (E.g., number of bedrooms), contained aggregated data (E.g., houses sold per year), or were targeted towards rental property availabilities and didn't contain sales prices.</li></ul>

### New Items

Item	Discussion Points
<b>Should we pivot to a different topic area?</b>  <i>Owner: All</i> <i>Deadline: n/a</i>	<ul style="list-style-type: none"><li>Team had exhausted all potential sources for appropriate housing data.</li><li>Discussion held around how we could make current data work. Ideas floated to pivot to different target for analysis (E.g., number of available houses instead of sales price). Decided that the sourced datasets could not offer a sufficiently detailed topic for analysis.</li><li>Discussion around which topic would be better. The team spent the next hour exploring datasets for each of the remaining topics, and was most impressed by the data available for <i>Air Quality and Health</i>.</li><li>Specifically, Seth identified two very strong sources of information for NSWs Health and Air Quality.</li><li>Explored both dataset sources in detail and felt satisfied that there a substantial amount of relevant and timely features to analyse.</li><li>Locked in new topic.</li></ul>
<b>Search for additional datasets.</b>  <i>Owner: Henry</i> <i>Deadline: Sept 17, 2024</i>	<ul style="list-style-type: none"><li>The team is to continue searching for relevant datasets over the next few days, to see if there is any more information we can integrate to deepen the analysis.</li></ul>
<b>Data Collection.</b>  <i>Owner: Matthew</i> <i>Deadline: Sept 17, 2024</i>	<ul style="list-style-type: none"><li>Relevant Air Quality Features to be selected, downloaded, and stored in shared repo.</li><li>Relevant Health Statistic datasets to be selected, downloaded, and stored in shared repo.</li></ul>

Action Items

Action Item	Owner	Deadline
• Download Health and Air Quality Datasets	Matthew	17 Sept, 2024
• Search for Additional Relevant Datasets	Henry & Seth	17 Sept, 2024

Next Meeting

Location	Swinburne Hawthorn Campus Room EN401
Date & Time	Tuesday, September 17, 2024 8:30am

*Henry Richardson*  
Henry Richardson  
*Meeting Facilitator*

*Matthew Cross*  
Matthew Cross  
*Secretary*

*Seth Kalantzis*  
Seth Kalantzis  
*Attendee*

# Minutes

## Machine Learning Model Development Phase

Meeting No. 9

Location	Swinburne Hawthorn Campus Room EN401
Date & Time	Tuesday, September 17, 2024 8:30am

Contributors	Henry Richardson Matthew Cross Seth Kalantzis	Meeting Facilitator Secretary
--------------	---	----------------------------------

### Action Items from Previous Meeting

Item	Discussion Points
<b>Download Health and Air Quality Datasets</b>  <i>Owner: Matthew</i> <i>Status: In-Progress</i>	<ul style="list-style-type: none"><li>All NSW Health data downloaded and ready for pre-processing.</li><li>Health Stat data was too substantial to download in a single file. The connection gateway to NSW Health Stats would time out before it could be downloaded.</li><li>Data downloaded in 4-year partial datasets and will need to be combined before, or as part of, pre-processing.</li></ul>
<b>Search for Additional Relevant Datasets</b>  <i>Owner: Henry</i> <i>Status: In-Progress</i>	<ul style="list-style-type: none"><li>No datasets identified that were of the same quality and from a reputable source.</li><li>Seth identified that we may be able to incorporate carbon / greenhouse gas emission data. More research required.</li></ul>

### New Items

Item	Discussion Points
<b>Begin Pre-Processing Data</b>  <i>Owner: All</i> <i>Status: In-Progress</i> <i>Deadline: Sept 20, 2024</i>	<ul style="list-style-type: none"><li>All members of the team to work on pre-processing the data individually before the next meeting. The team will review each-others progress and select the most appropriate approaches from across the implementations.</li><li>Each team member assigned a different dataset to focus on initially.</li></ul>
<b>Issues with GitHub</b>  <i>Owner: Matthew &amp; Seth</i> <i>Status: Completed</i>	<ul style="list-style-type: none"><li>Matthew unable to connect to team repo with his student account.</li><li>Seth added Matthew's personal account to the repo. Seemed to have resolved the issue.</li></ul>

Action Items

Action Item	Owner	Deadline
<ul style="list-style-type: none"><li>Pre-Process Assigned Dataset</li></ul>	All	Sept 20, 2024

Next Meeting

Location	Swinburne Hawthorn Campus Late Lab
Date & Time	Friday, September 20, 2024 3:30pm

  
Henry Richardson  
*Meeting Facilitator*

  
Matthew Cross  
*Secretary*

  
Seth Kalantzis  
*Attendee*



# Minutes

## Machine Learning Model Development Phase

Meeting No. 10

Location	Swinburne Hawthorn Campus The Node, Upstairs
Date & Time	Friday, September 20, 2024 3:30pm

Contributors	Henry Richardson Matthew Cross Seth Kalantzis	Meeting Facilitator Secretary
--------------	---	----------------------------------

### Action Items from Previous Meeting

Item	Discussion Points
<b>Pre-Process Assigned Dataset</b>  <i>Owner: All</i> <i>Status: In-Progress</i>	<ul style="list-style-type: none"><li>Attempts made by all to pre-process. Matthew had made good progress on Air Quality data, Henry and Seth had made good progress on Health Stat Data.</li></ul>

### New Items

Item	Discussion Points
<b>Issue Merging Air Quality and Health Data</b>  <i>Owner: Matthew</i> <i>Status: Completed</i>	<ul style="list-style-type: none"><li>Identified while pre-processing that the Local Health Districts in the Health Stat data and the Air Quality Collection Sites in the Air Quality Data were hard to map to one another.</li><li>Unable to identify a source online which provided a list of the suburbs within each local health district. We were, however, able to identify a Google Map which highlighted the specific regions that Local Health Districts covered.</li><li>Matthew created a secondary Google Map file, building upon the first, which dropped pins for each Air Quality Collection site within the relevant Local Health District. This data was exported and is being used as an additional dataset to merge the data the team currently has.</li></ul>
<b>Division of Machine Learning Tasks</b>  <i>Owner: Henry</i> <i>Status: Completed</i>	<ul style="list-style-type: none"><li>Data was now processed sufficiently to allow the team to begin implementing the machine learning models which they had been researching and building. Decision was made to distribute models based on the areas which the team had researched the most extensively.</li><li>Henry to complete linear regression models and analysis.</li><li>Seth to complete classification models and analysis.</li><li>Matthew to complete clustering models and analysis.</li></ul>

Action Items

Action Item	Owner	Deadline
• Continue to pre-process data	Matthew	Sept 24, 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 Room EN401
Date & Time	Tuesday, September 24, 2024 8:30am

*Henry Richardson*  
Henry Richardson  
Meeting Facilitator

*Matthew Cross*  
Matthew Cross  
Secretary

*Seth Kalantzis*  
Seth Kalantzis  
Attendee

# Minutes

## Machine Learning Model Development Phase

Meeting No. 11

Location	Swinburne Hawthorn Campus Room EN401
Date & Time	Tuesday, September 24, 2024 8:30am

Contributors	Henry Richardson Matthew Cross Seth Kalantzis	Meeting Facilitator Secretary
--------------	---	----------------------------------

### Action Items from Previous Meeting

Item	Discussion Points
<b>Continue to pre-process data.</b>  <i>Owner: Matthew</i> <i>Status: In-Progress</i>	<ul style="list-style-type: none"><li>Decision to normalization and splitting of data to be conducted independently with each individual model, to best fit the needs of the particular analysis.</li><li>All data cleaned, processed and integrated.</li><li>Additional work to standardise formatting of notebooks and comments.</li></ul>
<b>Complete the regression model</b>  <i>Owner: Henry</i> <i>Status: In-Progress</i>	<ul style="list-style-type: none"><li>Strong progress on linear regression model.</li><li>Identified that out initial hope of finding very strong correlations may have been too optimistic.</li></ul>
<b>Complete the classification model</b>  <i>Owner: Seth</i> <i>Status: In-Progress</i>	<ul style="list-style-type: none"><li>Classification proved challenging for the data collected for this assignment.</li><li>Focus was on linear regression and time series analysis during the week.</li><li>Will continue to work on classification model, specifically KNN.</li></ul>
<b>Complete the clustering model</b>  <i>Owner: Matthew</i> <i>Status: In-Progress</i>	<ul style="list-style-type: none"><li>Initial attempts made on K-means, DBSCAN and hierarchical clustering analysis.</li><li>More work to be done to compare metrics and determine most appropriate model.</li></ul>

### New Items

Item	Discussion Points
<b>File Paths Too Long</b>  <i>Owner: Henry</i> <i>Status: In-Progress</i>	<ul style="list-style-type: none"><li>Identified issues with pulling files from git as the file paths had become too long.</li><li>Seemed to only be an issue for Henry's machine. Matthew and Seth able to pull data successfully.</li><li>Henry downloading and uploading files manually as a temporary solution.</li></ul>
<b>Threshold for Strong Correlation</b>  <i>Owner: Henry</i> <i>Status: Complete</i>	<ul style="list-style-type: none"><li>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.</li></ul>

- 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*  
Henry Richardson  
*Meeting Facilitator*

*Matthew Cross*  
Matthew Cross  
*Secretary*

*Seth Kalantzis*  
Seth Kalantzis  
*Attendee*

# Minutes

## Machine Learning Model Development Phase

Meeting No. 12

Location	Swinburne Hawthorn Campus Library, Level 3
Date & Time	Friday, September 27, 2024 2:00pm

Contributors	Henry Richardson Matthew Cross Seth Kalantzis	Meeting Facilitator Secretary
--------------	---	----------------------------------

### Action Items from Previous Meeting

Item	Discussion Points
<b>Update comments in data cleaning files.</b>  <i>Owner: Matthew</i> <i>Status: Completed</i>	<ul style="list-style-type: none"><li>Comments and markdown updated within files.</li><li>File structure reworked to reduce file path length.</li></ul>
<b>Complete the regression model</b>  <i>Owner: Henry</i> <i>Status: Completed</i>	<ul style="list-style-type: none"><li>Linear Regression model completed.</li><li>Capable of analysing and comparing any relationship within the data.</li></ul>
<b>Complete the classification model</b>  <i>Owner: Seth</i> <i>Status: In-Progress</i>	<ul style="list-style-type: none"><li>KNN model in progress. Will continue to work on the analysis, but acknowledgement was made the classification models are not the most appropriate for the dataset.</li><li>Time series model and additional analysis of linear regression model completed by Seth.</li></ul>
<b>Complete the clustering model</b>  <i>Owner: Matthew</i> <i>Status: Completed</i>	<ul style="list-style-type: none"><li>K-Means analysis completed.</li><li>Additional metrics for analysis to be added to the notebook to enhance understanding of the models effectiveness.</li></ul>

### New Items

Item	Discussion Points
<b>Division of Final Tasks for Assignment 2</b>	<ul style="list-style-type: none"><li>All to tidy and upload their finalised machine learning models.</li><li>All to complete sections on the report relevant to the models that they developed.</li><li>Henry to handle report Introduction and Problem Statement sections.</li><li>Matthew to handle report Data Collection and Data Processing sections.</li><li>Seth to handle Conclusion section.</li><li>All to update README.md file with steps relevant to the models they completed.</li></ul>

Action Items

Action Item	Owner	Deadline
• Tidy analysis files and upload to GitHub	All	Sept 28, 2024
• Complete implementation section of report	All	Sept 28, 2024
• Complete Introduction and problem framing section of report	Henry	Sept 28, 2024
• Complete data collection and processing section of report	Matthew	Sept 28, 2024
• Complete conclusion of report	Seth	Sept 29, 2024
• Update README.md file	All	Sept 29, 2024
• Fill in contribution form	All	Sept 29, 2024
• Condense all meeting minutes to a single file	Matthew	Sept 29, 2024
• Submit all files and documents to canvas	Matthew	Sept 29, 2024

Next Meeting

Location	Swinburne Hawthorn Campus Room EN401
Date & Time	Tuesday, October 02, 2024 8:30am

*Henry Richardson*  
Henry Richardson  
*Meeting Facilitator*

*Matthew Cross*  
Matthew Cross  
*Secretary*

*Seth Kalantzis*  
Seth Kalantzis  
*Attendee*