

SIT723 – Handover Documentation

Student: Andrew Hallam – 220477645

Supervisors: Dr Atul Sajjanhar, Dr Glory Lee

Topic: Cognitive Workload Detection and Classification



Contents

Purpose of document	3
Copies of papers referenced in this research	3
Literature review.....	3
Research report	3
Data.....	3
Machine learning models and code.....	3
Code breakdown.....	4
Proposed steps for future	4

Research Handover

Purpose of document

The following document summarizes my research in a way that will allow for it to be continued without me should it be required. I am planning to continue this project in trimester 3 2023, but will submit this document as a requirement of the unit. A copy of this document will also be available within the Github repository.

Please note that as Turnitin will often remove hyperlinks I have opted to include the full link rather than have it as a hyperlink within a word.

Copies of papers referenced in this research

All papers referenced are saved in the Github folder titled "papers". A direct link to the folder is available here: <https://github.com/AndrewHallam1/SIT723/tree/main/Papers>

These files are all saved as PDFs.

Literature review

A copy of my literature review is also included in the Github repository. A direct link to the folder is available here: <https://github.com/AndrewHallam1/SIT723/tree/main/Literature%20Review>.

This document is saved in the Microsoft word format to allow for easier editing in future if required.

Research report

The research report is available as a pdf on the github in the following folder:

<https://github.com/AndrewHallam1/SIT723/tree/main/Research%20Report>

Please note that as this task is due prior to the finalization of the research report, it does not yet exist in the folder but will upon completion of the unit.

Data

Unfortunately, Github will only accept file sizes smaller than 250mb. As the size of the data exceeds that of the limit on Github (even when zipped), links to the data provided by the data authors can be found below:

COLET: <https://zenodo.org/record/6801166#.Y9pBXXYzblU>

MLA: <https://tinyurl.com/EyeTrackData>

Machine learning models and code

In this iteration 16 machine learning models were created, along with feature analysis and feature extraction. This has all been completed in a single python code.

All model code and code iterations are available on Github. A direct link is available here:

<https://github.com/AndrewHallam1/SIT723/blob/main/SIT723v1.ipynb>

Code breakdown

A brief summary of the functions of my code is below, however please note that I have made comments throughout the code in such a way that it explains what is happening at each step, as such, minimal explanation is included below:

1. Imports all required packages – you may need to install these if not already installed
2. Reads in all csv files in your working repository (ensure that you have set this and have all of the COLET data as csv files – this is how it differentiates between COLET and MLA datasets)
3. The csv files are appended together to create larger datasets
4. These datasets are then split into the required features – further details are included in the code comments
5. Following feature extraction the features are joined back together to create the final datasets
6. A min-max scaler is applied to the data
7. Feature analysis is conducted using linear regression and random forest
8. Feature analysis values are output as charts to better visualize them
9. Features are dropped manually following this feature analysis (this must be updated if you are going to change the dropped features)
10. Machine learning model hyperparameters are tuned and the optimal hyperparameters are output – these are to be used in the final machine learning models and are manually input
11. Machine learning models are created for multiclass
12. The dataset is merged on the predictor classes to create binary models (these are noted in detail in the code)
13. Binary machine learning models have their hyperparameters tuned
14. Finally, the binary models are run

Proposed steps for future

I have proposed multiple next steps for the future of the research:

- Inclusion of additional features
 - o Include the mean, median, mode, maximum and minimum of each of the predictors
- Inclusion of deep learning models
 - o I have not been decided upon which models to use yet and more research is required to identify these
- Inclusion of additional non-deep learning models including KNN and logistic regression
 - o This is to improve upon the existing models by broadening the scope. Only the best performing models as noted in previous research were included, however others may be useful for modelling the aforementioned data
- Multiple iterations of models to average test-train split results
- Oversampling to address data imbalance