

# Structure of Appendices:

- **2466057\_Figures\_and\_Tables.pdf**
  - Appendix A: Model Hyperparameters and Values
  - Appendix B: Model Configuration and Implementation
  - Appendix C: Comparison of Models Between Dummy Classifier
  - Appendix D: Correlation of All the Metrics
  - Appendix E: Duplicate for Figure 3 With Full Dataset
  - Appendix F: Duplicate for Figure 4 With Full Dataset
  - Appendix G: Difference in Performance Across Top Four Models - Across the Four Metrics
- **2466057\_Code**
  - Python Code
    - Metrics.py
    - Model.py
    - Pipeline.py
    - FeatureElimination.py
    - run\_pipeline.py
    - run\_rfe.py
  - R Code
    - tidy\_Data.R
    - setup\_subtypes.R
    - setup\_full.R
    - graph\_functions.R
    - process\_merged\_data\_functions.R
    - Hyperparameters\_Final\_Phase.Rmd (+ HTML)
    - Results\_Final.Rmd (+ HTML)
    - RFE\_Results.Rmd (+ HTML)
  - BASH Code
    - run\_pipeline.sh
    - concatenate.sh
    - activate\_env.sh
    - deactivate\_env.sh
    - test\_pipeline.sh
    - venv\_test.sh

The supplementary graphs and figures are grouped together in a single PDF file (2466057\_Figures\_and\_Tables.pdf) while all the code is merged into one folder (2466057\_Code). The code is then divided by the language - which also reflects the different functions of the code. Python is used for training the models, determining the hyperparameters and generating the metric data used in the visualisations. The parsing and visualisation of the data is done in R. Some of these are done with R scripts and some work is done in RMarkdown files, where the rendered HTML files are also provided. These HTML files might be easier to view as they work in any browser and hide the code required to generate the plots. Finally, since the model training is executed on a remote server, some BASH scripts are also supplied to illustrate some of the practical elements of this project, like using a virtual environment for Python packages.