**Planning section**

**Project initiation**

We formed a group and held our initial meeting on the 10th of November 2019. At this meeting we made a series of decisions for the project:

1. Assignment of sections to group members
2. Workflow methodology
3. Project principles
4. Agreed a work schedule
5. Established communication
6. Programming conventions
7. Collectively this was our first experience using machine learning, so rather than prioritizing experience, we discussed the aspects and requirements of each section. We then individually volunteered for sections based on personal preferences.
8. We identified that each of the data handling steps are sequential, requiring input from the preceding stage. For example, the input of the feature selection step is the pre-processed data. This constraint necessitates a waterfall-like workflow.
9. Due to time constraints and our relative inexperience, we determined that our project should favor speed and simplicity over complexity, for example using available libraries like scikit-learn rather than attempting to build our own tools from scratch.
10. Factoring 10 weeks to complete the project, we agreed to attempt to complete 2 full iterations prior to the Christmas break. The planned schedule is depicted in figure 1.
11. To enable distributed development of the project, we set up a WhatsApp group for communication and created a Github repository to store project files.
12. Further to point 5, we agreed that all programming should be performed in the Python programming language, and each section implemented as a function with documented inputs and outputs to allow easy use by the other group members.



Figure 1. Planned work schedule.

**Project review**

We held two review sessions, the first on the 12/12/2019 and a final review session on the 09/01/2020.

In the first session, we reported the conclusions of the first two iterations, and identified critical areas for improvement:

* Trying different scaling methods such as MinMax, StandardScaler
* Exploring dimensionality reduction techniques for feature selection like PCA
* Reviewing the algorithm selection, adding in K-nearest neighbours and Naïve Bayes Classifier
* Expanding the tuning parameters search size.
* Exploring further evaluation parameters such as F1 score, False Alarm rate, Michaels Correlation Coefficient

In the final review session, we discussed the conclusions of the project and assigned responsibilities for the planning section and the future of the report and code clean up.

**Summary**

* We delivered the project closely to the planned schedule, with several high-performance models developed by the Christmas break, allowing further research than anticipated.
* The WhatsApp group and Github repository were highly effective tools for sharing knowledge.
* By undertaking this project, we have all gained an insight into the application of machine learning algorithms to a real-world application.