**Preliminary Project Report Plan**

­­**Title:** Automated Feature and Hyperparameter Selection for Artificial Neural Networks

**Introduction**

* Describe problems with hyperparameter selection
  + The performance of a model is heavily dependent this
  + I came into this problem in neural networks module
  + Takes too long
  + Require experience
  + Random search takes a long time and needs a lot of computing power
  + Grid search is practically impossible in high dimensionality space
* Describe problems with feature selection
  + Lots of methods, option paralysis
  + Takes time to code methods
  + Provide a package to do it all for you
* Describe how I will use skills learnt in Neural Networks
  + Creating a NN from scratch helped me learn libraries
  + Ensemble methods
* Describe how I will use skills learnt in Data Mining
  + Feature selection methods (Forward pass, backwards pass, mutual information, genetic algorithms)
* Describe how I will use skills learnt in Machine Learning
  + Linear algebra with numpy
  + SKLearn

**Aims and Objectives**

* Can neural networks be reliably tuned using Bayesian optimisation?
  + Data from testing on multiple datasets
* Does Bayesian optimisation outperform Random search?
  + Data comparing tests performed on both, grid search as a control
* Does choosing predicted good starting parameters for Bayesian optimisation improve performance? (time, loss)
  + Random starting values vs chosen
* I will create a database containing optimal hyperparameters for select datasets along with meta features the describe the dataset
  + Help chose starting parameters for closely related datasets
  + Used to find out if choosing predicted starting parameters increases performance
* I plan to create a python package that will perform feature selection and hyperparameter tuning for artificial neural networks in classification tasks
  + I can use the package to perform tests and show the effectiveness of Bayesian optimisation

**Methods**

* Why Bayesian Optimisation?
* Why PyTorch?
  + Describe advantages over other libraries
* Why SMAC?
  + Describe advantages over other libraries
* Why a python package?
  + Integrated straight into code, can be used alongside other methods
* Performing random search to find optimal for database
  + Works well
  + Takes a long time but this is a onetime task
* Coding in python
  + Quick development time
  + Has all the necessary libraries

**Project Plan**

* Gantt Chart

**Planned Work**

* Extension of the Gantt chart