Algorithm Plan

Initialisation:

- Create a SQL database to store previous learnt hyperparameters and their datasets
- Find a range of classification problem datasets and use random search or grid search to find their optimal hyperparameters and store inside the database along with a number of meta features about the dataset.

Algorithm:

1. User uploads a CSV of data to be classified and inputs the name of column that is the label.
2. The program uses a feature selection technique (genetic algorithm, forward/backwards search, mutual information) to be decided to select the most important features.
3. Next the K (I will test different values) most similar datasets (found using a distance measure) hyperparameters will be fetched from the database.
4. K SMACs will be initialised with these parameters then Bayesian optimisation will run for N iterations.
5. The program will then return the hyperparameters that produce the lowest loss to the user.
6. The program will also return the accuracy, loss and F1 score for the best performing network and an ensemble of networks using all K hyperparameters.