Assignment 2

Preprocessing: For preprocessing I load the train and test dataset with the help of CSV reader and drop their id value column also split the labels from the training dataset.

In further steps of preprocessing, I used the Normalization technique to normalize the dataset.For feature selection, I used SelectKBest to reduce the no of features to the certain threshold that is mentioned in the assignment guideline. I also used ExtraTreeClassifier to select features on the bases of the entropy.

Methodology: I used GridSearchCV to find the best configuration of parameters for the input parameter model. After getting the optimal parameters, I used to model with this parameter configuration and find the cross-validation accuracy of the model. And in the last, I predict the test dataset and store it in a CSV file.

Analysis: Here we are getting a subtle change in the accuracy when we run the model again and again so this is not a big issue we take the optimal parameters at the instant when we get reasonable high score.