**Support File Group Project Step 2:**

**GitHub link:** [**https://github.com/ewang24/dataMiningProject**](https://github.com/ewang24/dataMiningProject)

The algorithms that we use are Decision Tree Classifier, Gradient Boosting Classifier, Random Forest, Logistic Regression, and K-Neighbors to evaluate the data. For each algorithm and their parameters we classify using Grid Search to determine the best parameters and best scores, then we print these values to reference for the rest of our project. We determine the best model using Grid-Search and Cross Validation. We evaluate the best model, Random Forest, using an ROC curve with micro/macro averaging. The visualized graphs we are using are a plot of the Confusion Matrix, Feature Importance, feature distribution, and a plot of the model accuracies. The most important feature for our dataset is y-edge.

Class: CMPT300

Date: 11-17-17

|  |  |  |  |
| --- | --- | --- | --- |
| Member Name | Contribution (%) | How many meetings did he/she miss? (m/n) | Role |
| Zach | 25% | N | Researcher, helped visualize data and set objectives |
| Serena | 25% | N | Researcher, helped implement algorithms and troubleshoot |
| Logan | 25% | N | Researcher, helped visualize data, troubleshooting |
| Evan | 25% | N | Researcher, Team Leader, |