

Progress Report

Name: Anowarul Kabir

Date: 8th October, 2019

Data distribution:

Number of attributes: 16

Train set: 7494

Test set: 3498

Missing rows/columns: No

Data cleaning: No

System architecture:

Operating system: Ubuntu 18.04

Language and library used: Python 3.6, scikit-learn v0.21.3 DecisionTreeClassifier

Decision Tree(with below mentioned model)		Accuracy	Precision	Recall
Gini	Train set	1.0	1.0	1.0
	Test set	0.919	0.921	0.919
Entropy	Train set	1.0	1.0	1.0
	Test set	0.917	0.919	0.917

Table 1: Accuracy, Precision, Recall for different Decision Tree Model

Decision Tree(with below mentioned model)	1 st fold	2 nd fold	3 rd fold	4 th fold	5 th fold
Gini	0.956	0.964	0.958	0.951	0.965
Entropy	0.959	0.962	0.966	0.965	0.963

Table 2: 5-fold Cross Validation Score

The code can be found in the following github link.

https://github.com/akabiraka/cs584_data_mining/tree/master/hw1

In the next section, I have added confusion matrix of corresponding entropy and gini model for Decision tree classifier. The confusion matrix describes for which class the classifier is predicting wrong class label.



