

Assignment 1  
Decision Tree and AdaBoost for Classification

Rituparna Datta  
1505091

March 2020

# 1 Decision Tree

## 1.1 Performance Measure of Telco Dataset

Performance Measure	Training	Test
Accuracy	95.8126330731015%	74.02413058907025%
True positive rate (sensitivity, recall, hit rate)	88.47184986595174%	54.59317585301837%
True negative rate (specificity)	98.45559845559846%	81.22568093385215%
Positive predictive value(precision)	95.37572254335261%	51.87032418952619%
False Discovery Rate	4.6242774566473965%	48.12967581047381%
F1 score	91.7941585535466%	53.196930946291566%

## 1.2 Performance Measure of Adult Dataset

Performance Measure	Training	Test
Accuracy	90.60424966799468%	78.66533864541833%
True positive rate (sensitivity, recall, hit rate)	73.70967741935483%	57.86486486486486%
True negative rate (specificity)	96.14638447971782%	85.44014084507042%
Positive predictive value (precision)	86.25353884869456%	56.41633728590251%
False Discovery Rate	13.746461151305445%	43.58366271409749%
F1 score	79.48978112769967%	57.1314209472982%

## 1.3 Performance Measure of Credit Card Fraud Dataset

Performance Measure	Training	Test
Accuracy	99.96664442961975%	99.8806221691654%
True positive rate (sensitivity, recall, hit rate)	98.08917197452229%	44.0%
True negative rate (specificity)	99.97183346536397%	99.95429535746304%
Positive predictive value (precision)	90.58823529411765%	55.932203389830505%
False Discovery Rate	9.411764705882353%	44.067796610169495%
F1 score	94.18960244648318%	49.25373134328358%

## 2 Adaboost Implementation

### 2.1 Accuracy of Telco Dataset

Number of boosting rounds	Training	Test
5	79.09123180688677%	79.4180269694819%
10	79.3574724884629%	78.92122072391767%
15	79.26872559460418%	79.13413768630234%
20	80.36918707845226%	80.19872249822569%

### 2.2 Accuracy of Adult Dataset

Number of boosting rounds	Training	Test
5	82.3486506199854%	82.13811420982736%
10	83.2040315628937%	82.98140770252324%
15	83.61846031430277%	83.60039753912001%
20	83.7212386446522%	83.65869853917663%

### 2.3 Accuracy of Credit Card Fraud dataset (Using 20000 data)

Number of boosting rounds	Training	Test
5	99.80624999999999%	99.675%
10	99.8125%	99.775%
15	99.8125%	99.8%
20	99.825%	99.925%