

# **CSE472 (Machine Learning Sessional)**

## **Assignment 2: Logistic Regression and AdaBoost for Classification**



**Report Submitted By:**  
**1805060 - Shamim Al-Mamun**      **Section: A2**

### **How to run:**

To train the model and evaluate performance for both Logistic Regression and AdaBoost, on dataset

- Telco-customer-churn:
  - > Comment out lines 325 & 326
- Adult:
  - > Comment out lines 324 & 326
- CreditCard:
  - > Comment out lines 324 & 325

And then run “python3 1805060.py” from commandline.

## Performance Measure for Logistic Regression

### ☐ telco-customer-churn

Performance measure	Training	Test
Accuracy	73.33	73.77
True positive rate (sensitivity, recall)	0.99	0.99
True negative rate (specificity)	0.01	0.01
Positive predictive value (precision)	0.73	0.73
False discovery rate	0.26	0.25
F1 score	0.84	0.84

### ☐ adult

Performance measure	Training	Test
Accuracy	75.29	74.88
True positive rate (sensitivity, recall)	0.98	0.98
True negative rate (specificity)	0.01	0.02
Positive predictive value (precision)	0.75	0.74
False discovery rate	0.24	0.25
F1 score	0.85	0.85

□ credit-card

Performance measure	Training	Test
Accuracy	98.084	97.876
True positive rate (sensitivity, recall)	0.99	0.99
True negative rate (specificity)	0.37	0.28
Positive predictive value (precision)	0.98	0.98
False discovery rate	0.01	0.01
F1 score	0.99	0.99

**Accuracy of Adaboost with Logistic Regression**

➤ **Telco:**

Number of boosting rounds	Training	Test
5	73.33	73.77
10	73.33	73.77
15	73.33	73.77
20	73.33	73.77

➤ **Adult:**

<b>Number of boosting rounds</b>	<b>Training</b>	<b>Test</b>
<b>5</b>	75.29	74.88
<b>10</b>	75.29	74.88
<b>15</b>	75.29	74.88
<b>20</b>	75.29	74.88

➤ **Credit Card:**

<b>Number of boosting rounds</b>	<b>Training</b>	<b>Test</b>
<b>5</b>	98.08	97.87
<b>10</b>	98.07	97.85
<b>15</b>	98.08	97.87
<b>20</b>	98.08	97.87

➔ **Observations**

- ◆ Bining on numerical valued columns improve performance significantly.
- ◆ Columns having distributions biased to one value like 'native-country' in adult.csv(USA) are better to drop.
- ◆ Adaboost couldn't give significant performance improvement.