Assignment 1

How to run script:
Command in terminal:
- python3 1605071.py

The console asks for inputs
- Please enter dataset no: (1 for dataset 1 , 2 for dataset 2 , 3 for dataset 3)
- Please enter which type of model: (1 to run logistic regression , 2 to run adaboost)
- (If **adaboost** is selected as input)
    - Please enter the number of features to be selected through information gain:

Performance evaluation:
Logistic:
Dataset 1:

Performance measure	Training	Test
Accuracy	80.74192403265886	80.34066713981547
True positive rate (sensitivity, recall, hit rate)	54.10821643286573	53.76344086021505
True negative rate (specificity)	90.13778100072516	89.6817743490839
Positive predictive value (precision)	66.50246305418719	65.14657980456026
False discovery rate	33.497536945812804	34.85342019543974
F1 score	0.5966850828729282	0.589101620029455

Dataset 2:

Performance measure	Training	Test
Accuracy	85.02502994379779	85.06234260794791
True positive rate (sensitivity, recall, hit rate)	57.92628491263869	57.46229849193968
True negative rate (specificity)	93.6205501618123	93.59871330920788
Positive predictive value (precision)	74.22781500245138	73.51962741184298
False discovery rate	25.77218499754862	26.48037258815702

F1 score	0.6507163323782235	0.645067133683596

Dataset 3:

Performance measure	Training	Test
Accuracy	99.93811582435428	99.9420666409185
True positive rate (sensitivity, recall, hit rate)	77.51937984496125	78.0952380952381
True negative rate (specificity)	99.97625935337513	99.98241201611059
Positive predictive value (precision)	84.7457627118644	89.13043478260869
False discovery rate	15.254237288135593	10.869565217391305
F1 score	0.8097165991902834	0.83248730964467

Adaboost:
Dataset 1:

Number of boosting rounds	Training	test
5	80.58217962371317	80.69552874378992
10	80.45793397231097	80.62455642299503
15	80.20944266950657	82.68275372604684
20	81.11466098686546	78.99219304471256

Dataset 2:

Number of boosting rounds	Training	test
5	84.99431835631583	85.07462686567165
10	84.96053561008569	85.09305325225723
15	84.93903749884831	85.09305325225723
20	84.94210865759652	85.01934770591487

Dataset 3:

Number of boosting rounds	Training	test

5	99.93679913976607	99.9420666409185
10	99.9446992472953	99.93153330290369
15	99.93416577058966	99.93504441557529
20	99.9372380346288	99.95962220427653

Observations:
- Dataset 3 has too many negative samples; hence the amount of true positive is little.