

PRML

Lab-1

Given the binary classification problem, we used two models, logistic regression and random forests to make predictions. Finally these predictions are converted into labels using different thresholds.

Comparison:

Threshold: 0.5

Logistic Regression	Random Forest
TP : 4279 TN : 5425 FP : 2454 FN : 3600	TP : 5047 TN : 5519 FP : 2360 FN : 2832
Accuracy : 61.58141896179718	Accuracy : 67.05165630156111
Precision : 63.55265112134264	Precision : 68.1382476036182
Recall : 54.30892245208783	Recall : 64.05635232897576
F1 score : 0.5856830002737476	F1 score : 0.660342797330891

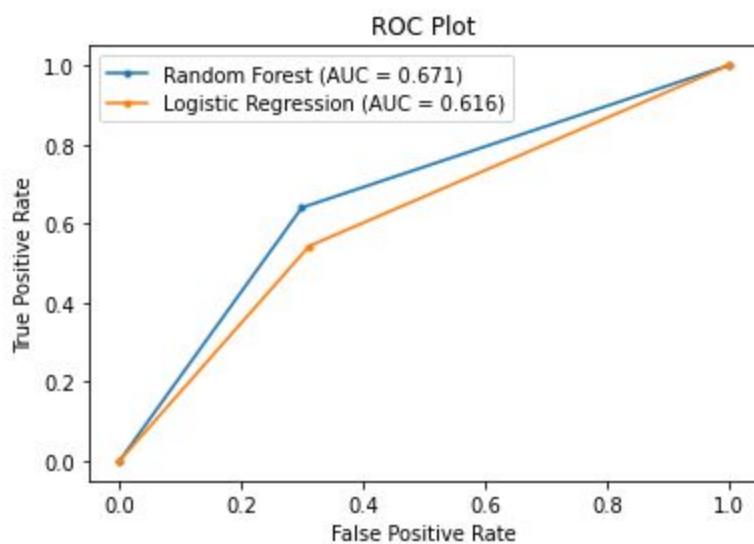
Threshold: 0.4

Logistic Regression	Random Forest
TP : 6535 TN : 2862 FP : 5017 FN : 1344	TP : 7411 TN : 1930 FP : 5949 FN : 468
Accuracy : 59.633202183018156	Accuracy : 59.27782713542328
Precision : 56.570290858725755	Precision : 55.471556886227546
Recall : 82.94199771544612	Recall : 94.06015991877142
F1 score : 0.6726365086717101	F1 score : 0.6978671312208673

Threshold: 0.6

Logistic Regression	Random Forest
TP : 2406 TN : 6858 FP : 1021 FN : 5473	TP : 2239 TN : 7417 FP : 462 FN : 5640
Accuracy : 58.789186444980324	Accuracy : 61.276811778144435
Precision : 70.20717829004961	Precision : 82.89522399111439
Recall : 30.536870161187966	Recall : 28.417311841604263
F1 score : 0.4256147178489298	F1 score : 0.4232514177693762

ROC curve :



From AUC values, we can say that Random Forest works better than Logistic Regression for this classification problem. This is due to the fact that higher the AUC, higher is the TPR for a given FPR.