

Logistic Regression v/s Random Forest

For the threshold of 0.5. we got:-

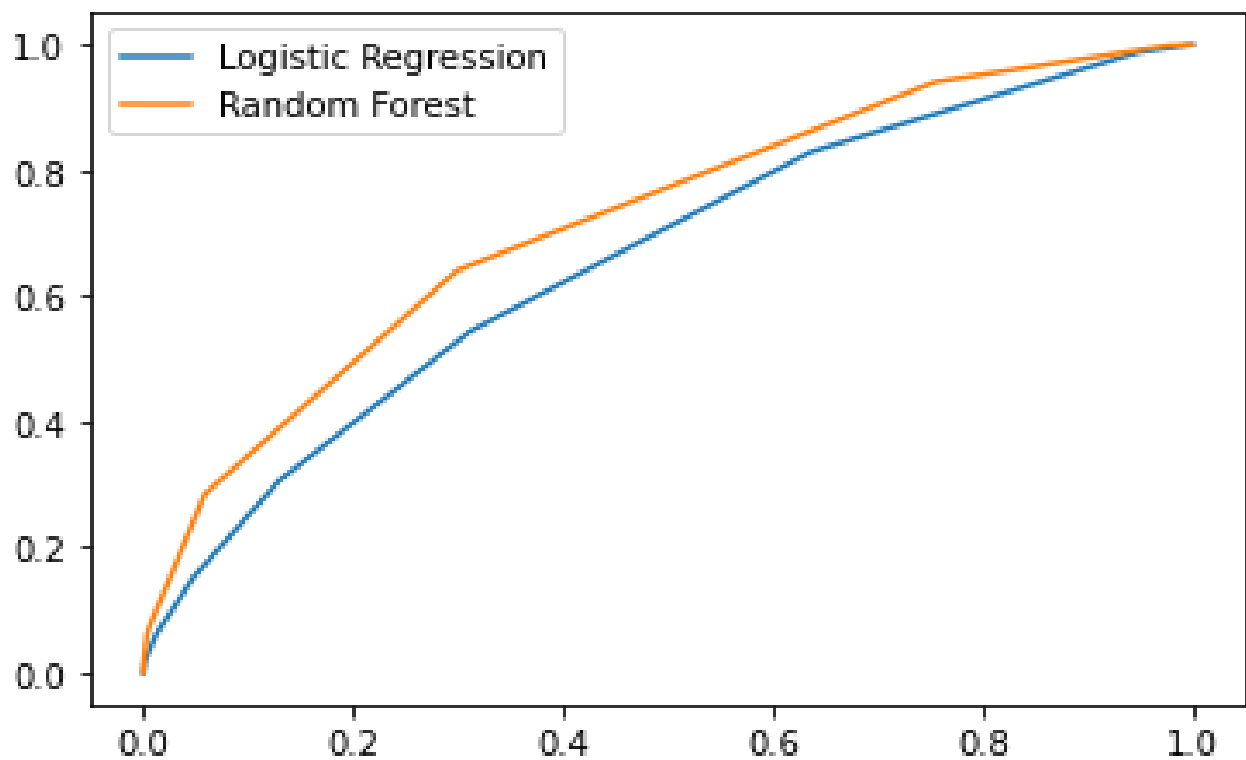
Metrics	Logistic Regression	Random Forest
True Positives	4279	5047
True Negatives	2862	5519
False Positives	5017	2360
False Negatives	1344	2832
Accuracy	61.581 %	67.051 %
Precision	63.552 %	68.138 %
Recall	54.308 %	64.056 %

Conclusion:-

While tuning these Models with different values of threshold, some of the insights brought up are as follows:-

1. Random Forest was always performing better than Logistic Regression, when the focus is on bringing out True Positives/Negatives and lessening the amount of False Classifications.
2. Metrics like Accuracy, Precision and Recall, brought up the picture about the prediction distribution and to how much extent the relevant information was saved, according to the dataset provided.

The ROC Plot for both models are as follows:-



X-axis:- False Positive Rate, Y-axis:- True Positive Rate