

I would consider the accuracy of approximately 97% (97,5 if calculated manually) to be quite good. However, setting all yhats to zero would result in an accuracy of about 90% because of the structure of the dataset, so I can’t really tell if the accuracy is good enough for drawing any conclusions from it.

The algorithm for calculating TPF/RPF turned out to be quite slow due to all the excessive looping and rather large dataset. I suppose a faster would be to sort the predictions (yhat) and use some sort of acumulative sum to only loop through the thresholds once, instead of looping every threshold/yhat for every possbile combination of other yhats. However, this would not change the result, only make the algorithm faster.