

The purpose of this analysis is to help the company to reduce losses and have more accurate prediction about the credit approval.

On the loan status we have 75036 rows of 0 and 2500 rows of 1. I need to work with the 2500 rows to predict the false negatives, it's a small group to work with, I needed then recall as a metric.

Linear model is pretty standard because there was no overfitting test and training and testing are almost the same result.

Best algorithm was in a vacuum even though there was imbalance, 11 out of 620 was incorrect which is 98% recall. That 11 negative score was the perfect curve