**Credit One Project**

**Report**

**By**

**Roslyn Badilla**



**March 2020**

**Final Report**

# Overview

Over the past year or so Credit One has seen an increase in the number of customers who have defaulted on loans they have secured from various partners, and Credit One, as their credit scoring service, could risk losing business if the problem is not solved right away. They have enlisted the help of our Data Science team to design and implement a creative, empirically sound solution. It is very important that we all understand from the start that this is not a typical data analytics problem as we have been given full authority to solve this problem with whatever tools and methods we need.

# Results

After training and testing a Random Forest model with client’s real data, the model could predict 81% of the cases successfully, but this is still a low metric in Bank terms.

As you can see in the chart below, we couldn’t predict 44% of Defaulted clients with only the data provided, so it will be better to have more data and features to create a better model.

However, the most important part of this investigation is the importance of the predictors. In the chart below are the featurews by importance, the first three show that the bill amount is highly correlated with the people defaulted, so it will be a good idea to check another behaviors related with buying, job stability or health problems.

And the limit of the credit and the amount to pay must be higher than past months to the point the client must default.

# Methodology

To achieve the results, the data was prepared, analysed and grouped, to obtain better results. The features deleted were the ones with high correlations between them. Two datasets were created with the difference that one of them had the limit\_bal variable binned.

We trained three classifiers and selected one to start a parameter tuning.

|  |  |  |  |
| --- | --- | --- | --- |
| **Classifier** | **Metric** | **X1** | **X2** |
| Random Forest | Score | 0.97881 | 0.974048 |
| Ada Boost | Score | 0.818905 | 0.817952 |
| Gradient Boost | Score | 0.827571 | 0.827286 |
| Random Forest Tuned | Score | 0.997571 |  |

The final results were:

* **Accuracy = 0.812**
* **F1\_score = 0.447**