

Machine Learning Outcomes – Bank Loan Analysis

Executive Summary Report

Prepared by Robes Fokoueng – May 2024

- Project Overview:** In this part of the project, we want to build a machine learning model to predict whether a customer will be a good or bad borrowers.
- Problem:** Since there are some customers who fail to pay their loans, we are seeking a way to identify this kind of customer and take proactive measures.
- Solution:** We built two tree-based classification models (Random Forest and XGBOOST). Both models were used to predict on a validation and test dataset, and final model selection was determined by the model with the best recall score.

Details

- The recall is ~ 0.046 greater than Random Forest model. Both models are not acceptable, but XGBoost model is the champion.
- This model correctly identified only 5.44% of the customers who are actually bad borrowers.

	model	precision	recall	F1	accuracy
0	RF CV	0.300318	0.012423	0.023811	0.859397
0	RF test	0.310345	0.008435	0.016423	0.860290
0	XGB CV	0.286900	0.065399	0.106392	0.848412
0	XGB test	0.280193	0.054358	0.091052	0.849922

Different scores for random forest and XGBoost.

Future suggestions

- Collect/add data, including credit score of each customer.

Results Summary

The resulting algorithm isn't usable to predict customers who might be bad borrowers.

Next Steps

As a next step, we wouldn't recommend using that model to predict customers who might be bad borrowers. However, additional data would be needed to realize significant improvement to the model.