



Project 3

Good or Bad Loan?

Leo Liu

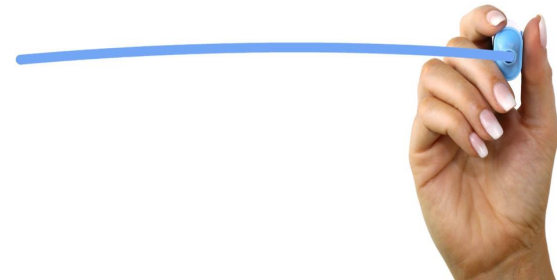
Objective

Detect risky clients



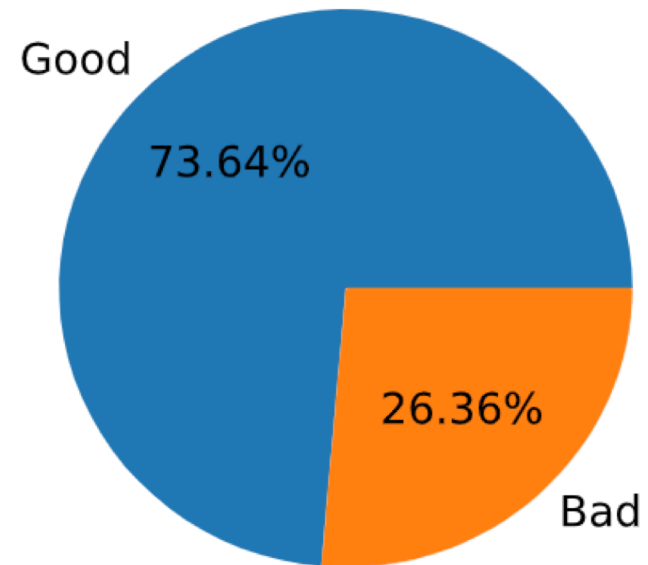
Improve policies

POLICIES



Data Source

- **Website:** [Kaggle.com](https://www.kaggle.com)
- **Overview:** Loan data through the 2007-2015
- **Dimensions:** 75 variables, 890k observations
- **Features:** Employment, Income, Homeownership, Credit Scores, Number of Financial Inquiries, Collection among Others.



 **LendingClub**

Model

Gradient Boosting Classifier

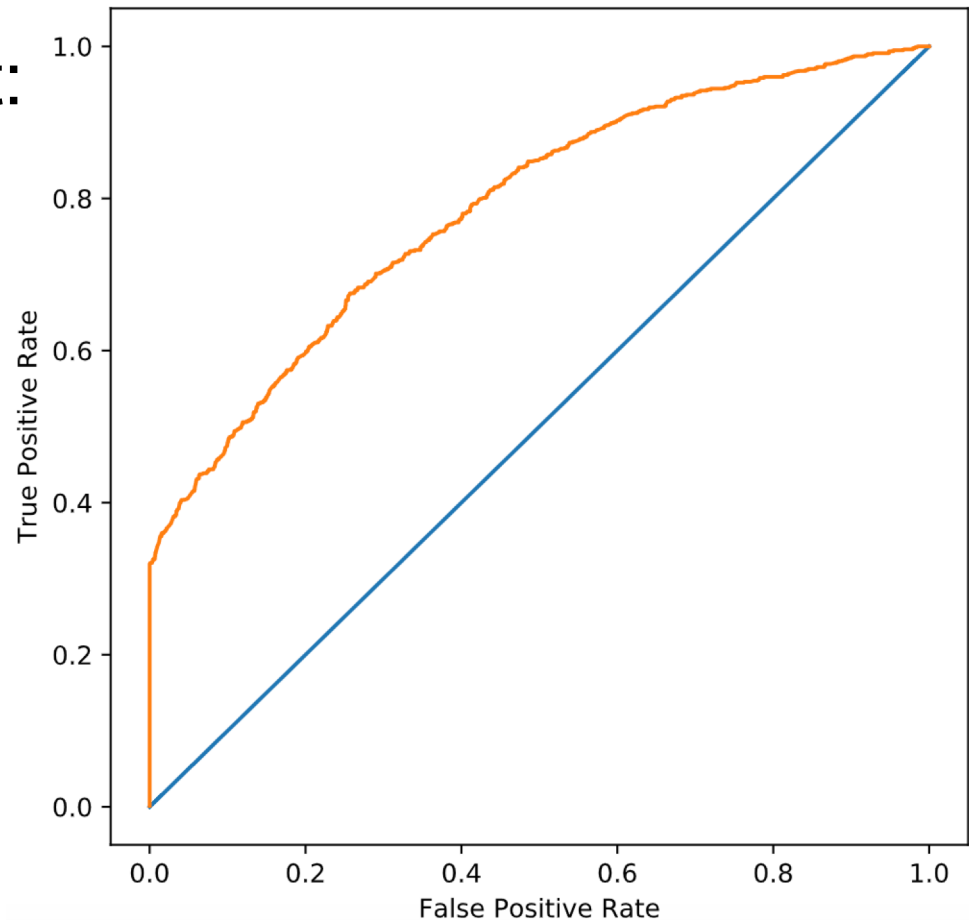
Classification Report:

Precision: 0.37

Recall: 0.86

f1-score: 0.51

ROC_AUC: 0.79



Model

Cost Benefit Analysis

Best Cutoff:

0.38

False Positive Loss:

\$1,902

False Negative Loss:

\$0

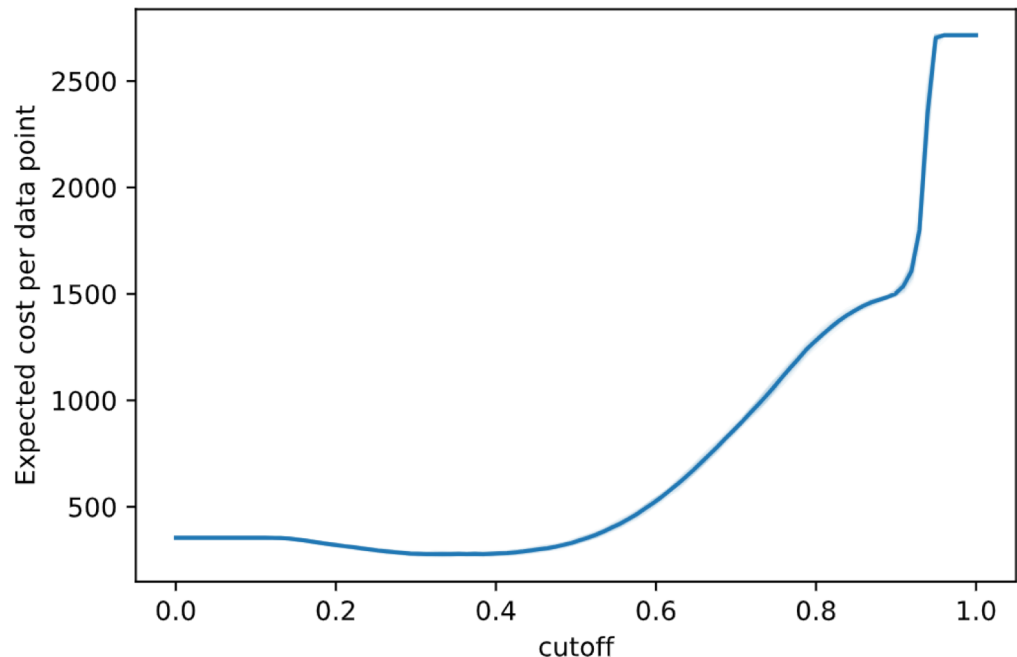
True Positive Gain:

\$8,509

Total Savings:

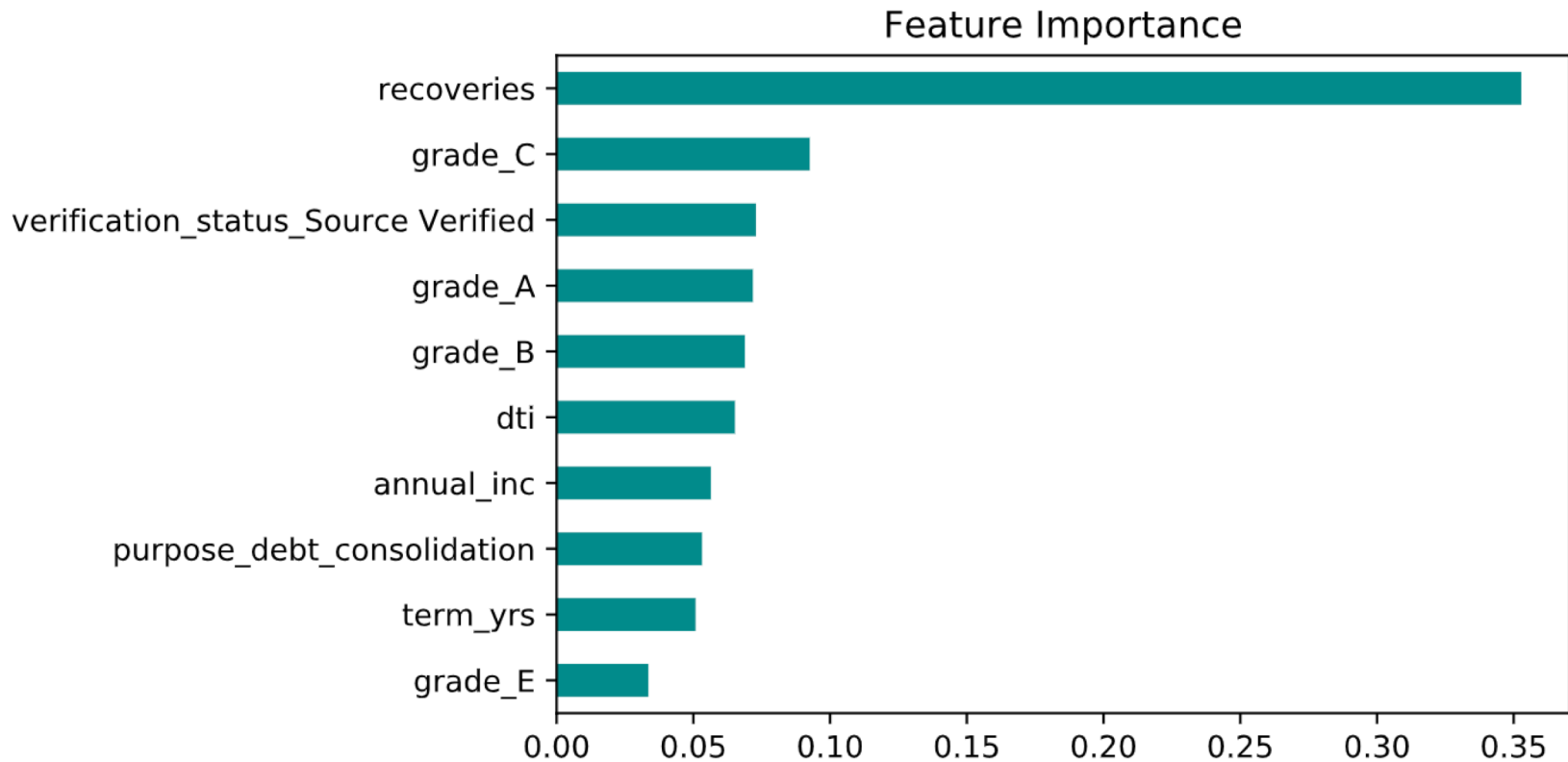
\$858/case

\$20M annually



Model

Feature Importance Analysis



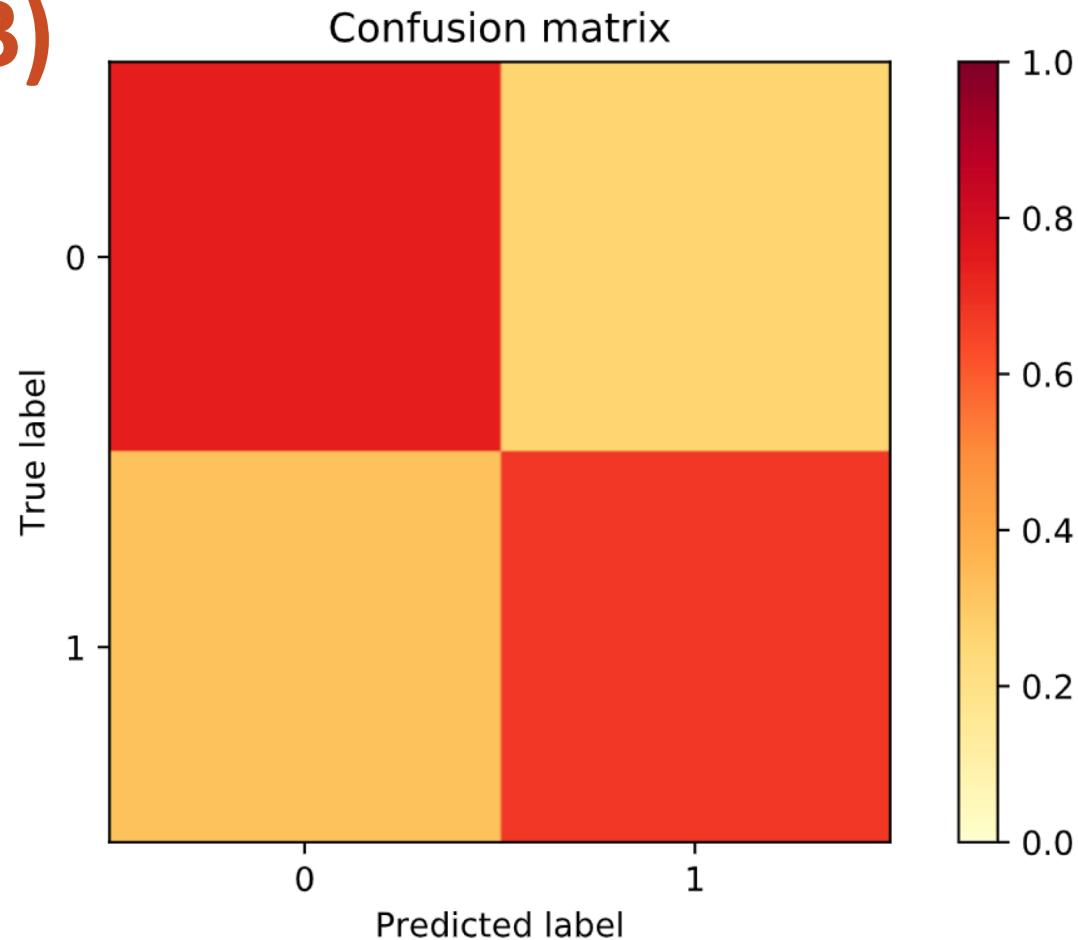
Future Work

- **Make use of unlabeled rows**
- **Design an application with flask**
- **Analyze bad loan data time wisely**
- **Incorporate more features if available**
- **Run larger scale dataset on AWS machines**

Thank You!

Appendix

Normalized Confusion Matrix(GB)



Appendix

Label Data Selection

