LOAN ELIGIBILITY PREDICTIONS

VIRTUAL INTERNSHIP EXPERIENCE

DATA SCIENCE

id/x partners

Erdin Nurjaman

Outline

- Introduction
- Goal
- Business Approach
- Understanding & Processing Data
- Modeling
- Conclusion

id/x partners

Introduction

- The financial industry needs to maintain the smooth flow of money traffic for the sake of its business continuity.
- In money loan services, bad debts can threaten the course of operational activities due to the absence of cash inflows.
- One of the reasons is the improper determination of potential borrowers so that it aborts potential creditors and instead passes potential creditors who have problems.

Goal

After looking at the previous problems, we plan to improve the system for selecting prospective creditors with two objectives:

- Streamlining the traffic of money coming out and in the company.

- Helping prospective borrowers who are responsible, so that they can take advantage of loans for consumption and business needs

Business Approach

We will utilize machine learning technology in determining potential borrowers who are worthy and who are not worthy of being given a loan.

Some of the advantages that can be felt from the use of machine learning technology include:

1. Speed in deciding thanks to the power of the computer.
2. Knowing the pattern or characteristics of the borrower.

Business Metrics

which has a significant effect

OCF

Operating Cash Flow Ratio

Cash for operating expenses will be available more often.

ROA

Return on Assets

The growth of assets is gradually increasing.

Understanding & Processing Data

The data that we will use is loan data provided by the company from 2007 to 2014 with an observation amount of about 200 thousand (after processing).

Each borrower is labeled "1" for those who successfully repay and labeled "0" for those who are late or fail to repay

Characteristics of Borrowers

LOAN STATUS

The loan status is owned by the borrower and is the information that we will predict.

TOTAL LOANS

The amount of the loan received by the debtor.

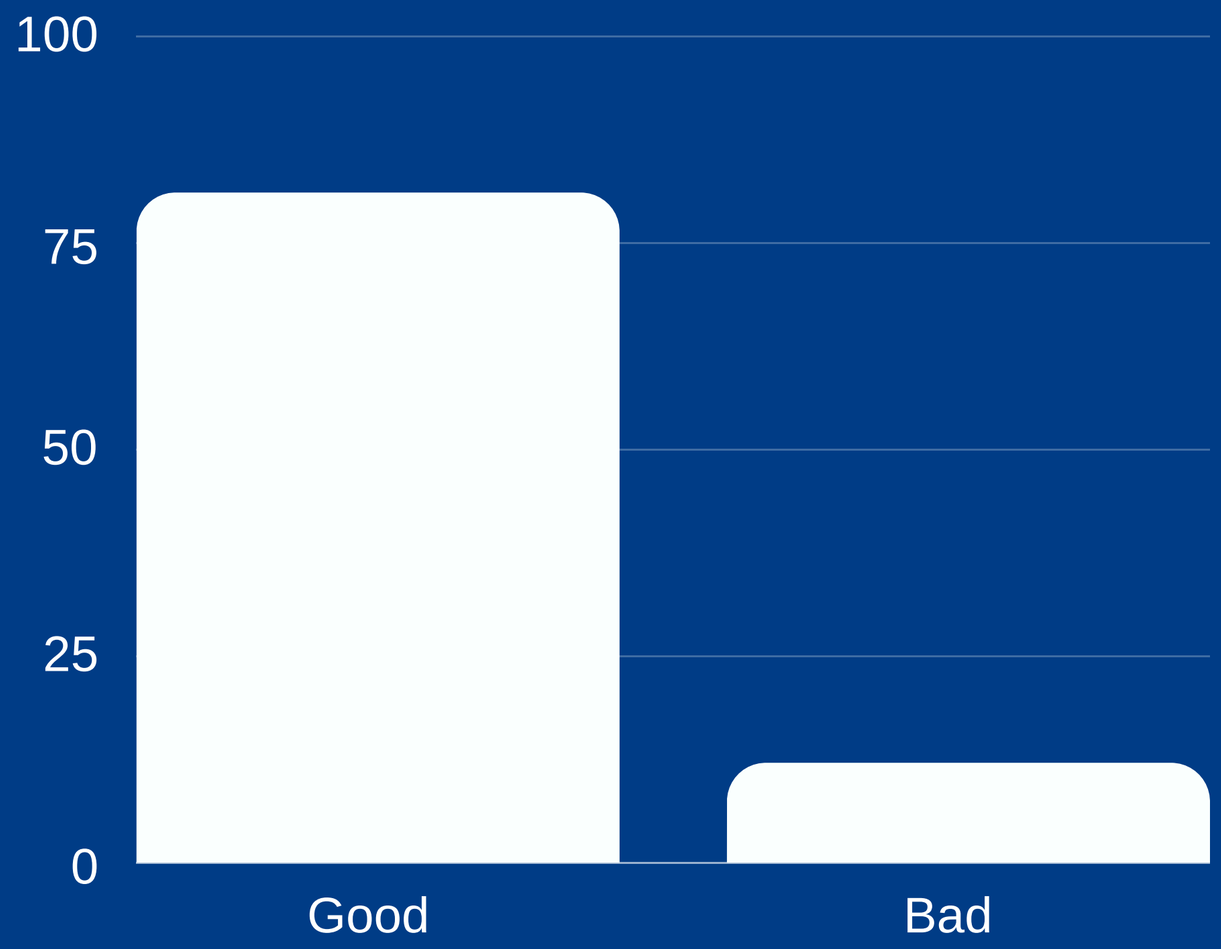
PURPOSE OF BORROWING

The reason why the debtor borrows money from the company.

LOAN RATING

The loan rating, which is an alphabetical letter of A - G, the closer to G, the greater the interest value.

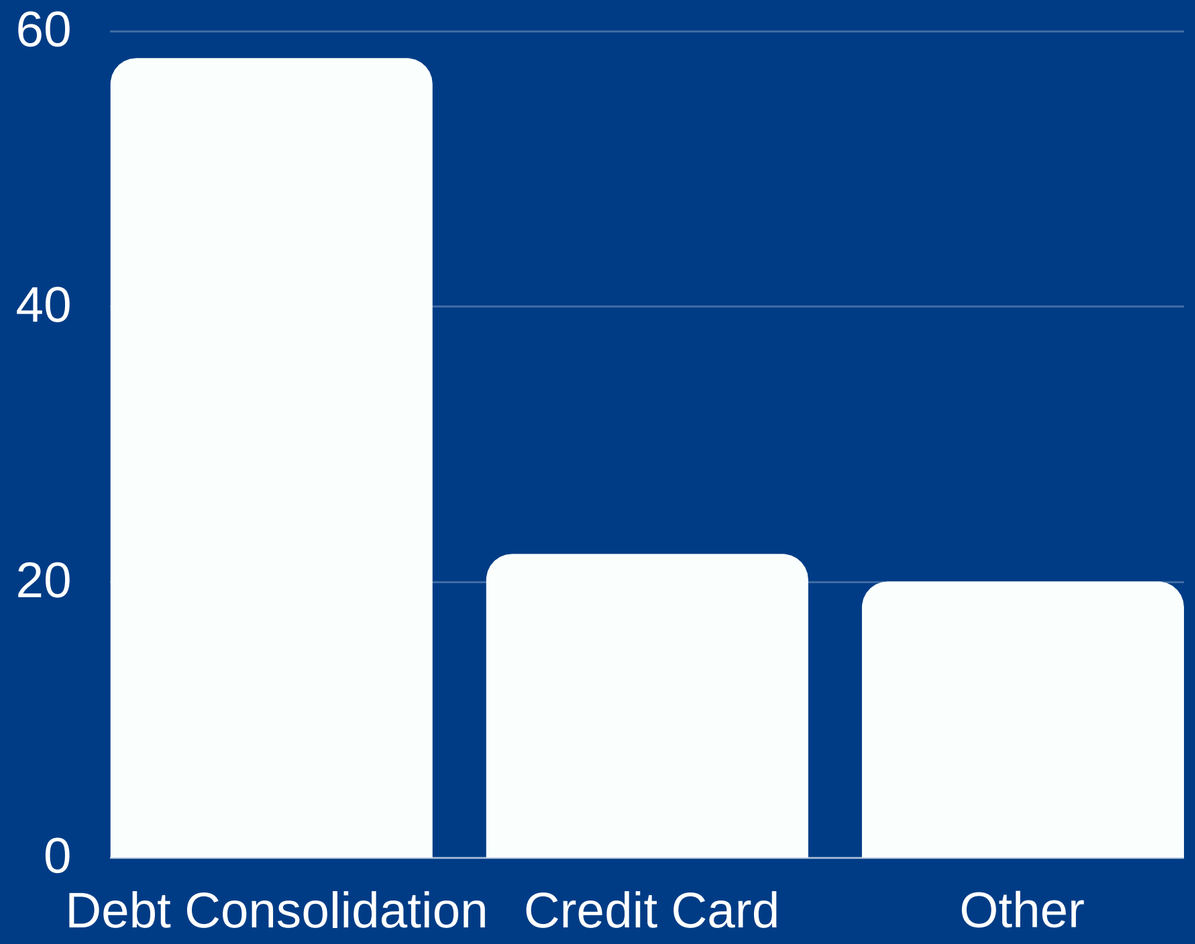
Loan Status



There are 81% of loans with good status, exceeding half of bad status loans.

id/x partners

Purpose of Borrowing



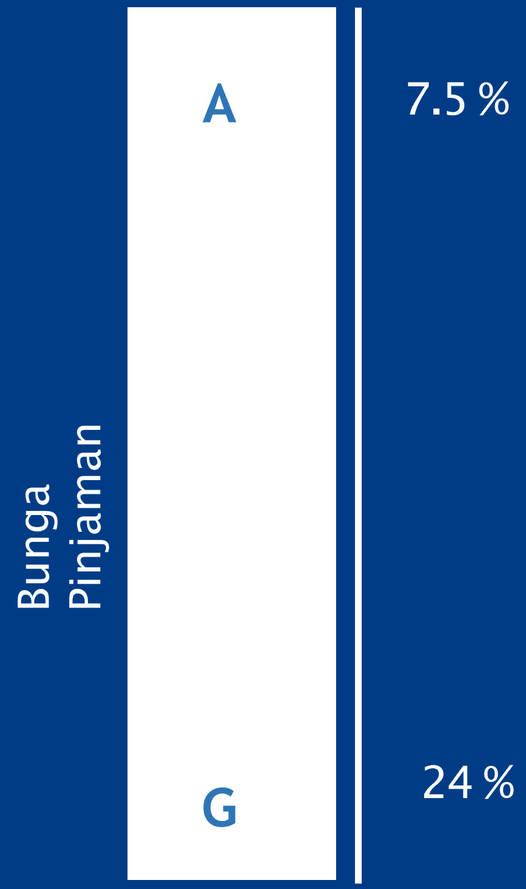
Most of the loans are made to patch up previous debts.

Total Loans

Loan Status	Mean	Std
Good	13,214	7,949
Bad	14,250	8,362

The average and deposit value of the total loan on a bad loan status is greater than the status of a good loan.

Peringkat Pinjaman



Each loan rating has its own interest value. Starting from A-rank with an average interest rate of 7.5% to G-rank with an average interest rate of 24%

id/x partners

Data Processing

The use of machine learning technology can be done when the data format we have is numerical, which can be processed by a computer. Some of the processing steps we do include:

1. Populate the missing value using mean, median or mode based on the distribution of data.
2. Normalize numerical data with StandardScaler.
3. Create a dummy variable on the rest of the categorical column.

Modeling

ALGORITHMS FOR PREDICTION

FORMS OF TRAINING

We separate the loan data into two parts, namely training and testing. The train data we will use to train the model and the test data we will use to test the model.

MODEL TYPES

We chose logistic regression to predict 2 categories of loan status.

Modeling

ALGORITHMS FOR PREDICTION

ADJUSTMENT

Given that the ratio between the categories of good and bad loan status is unbalanced, which is 81: 19, then we give a little weight to the good category and a large weight to the bad category.

PREDICTION RESULT

PREDICTION ACCURACY

96.7% ON TRAINING DATA

96.6% ON TEST DATA

The model can predict the loan status of customers precisely with probability of more than 90%

PREDICTED PERFORMANCE

PRECISION

88%

Possibility to predict bad loan status as bad.

RECALL

97%

The possibility not to incorrectly predict the status of a bad loan as good.

F1-SCORE

92%

A combination of precision and recall, it is likely how consistent the model is in predicting loan status.

id/x partners

ANALYSIS OF PREDICTED RESULTS

The logistic regression model has read the pattern of the characteristics or features of the borrower well. If we use this model, our probability of mischaracterizing a bad customer as good is 97% so that the risk of delay or default can be minimized.

Conclusion

BUSINESS BENEFITS

Judging from the prediction results, the use of machine learning models can improve OCF and ROA metrics.

CHARACTERISTIC ANALYSIS

Customers with a bad loan status tend to make loans with a larger amount than customers with good loan status.

PREDICTION RESULT

Models can recognize customers with bad status precisely.

id/x partners

THANK YOU!

LINKED IN

https://www.linkedin.com/in/erdinnurjaman

INSTAGRAM

https://www.instagram.com/erdinnurjaman