HOME CREDIT SCORECARD MODEL

Aditya Virgiansyah

PROBLEMS

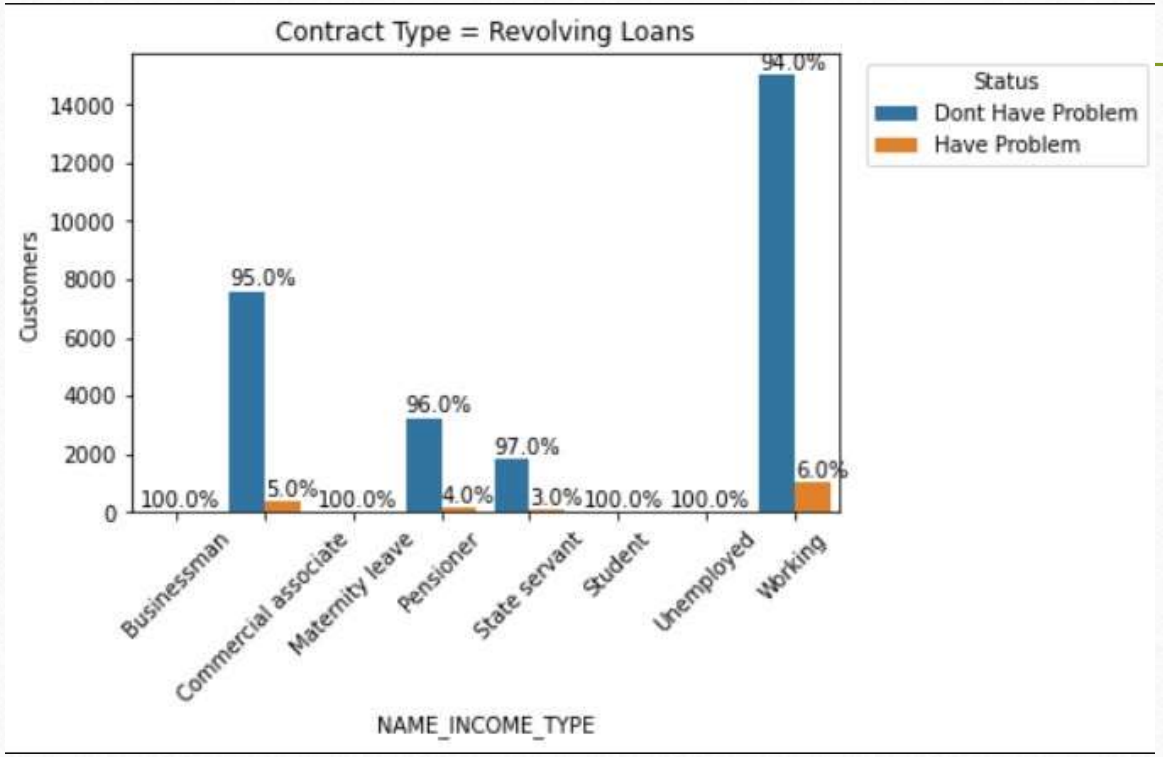
1. Home Credit Indonesia wants to create machine learning to help the team determine whether the loan applications from customers will experience problems in the credit repayment process or not.
2. From the existing data, Home Credit Indonesia wants to find out what customer criteria are that have not problems in the credit repayment process to help increase revenue.

DATASET

Data is provided by the Home Credit Indonesia team, The data consist of 122 columns and 307k rows containing customer data that has problems or not to paying off loans that have been given by Home Credit Indonesia.

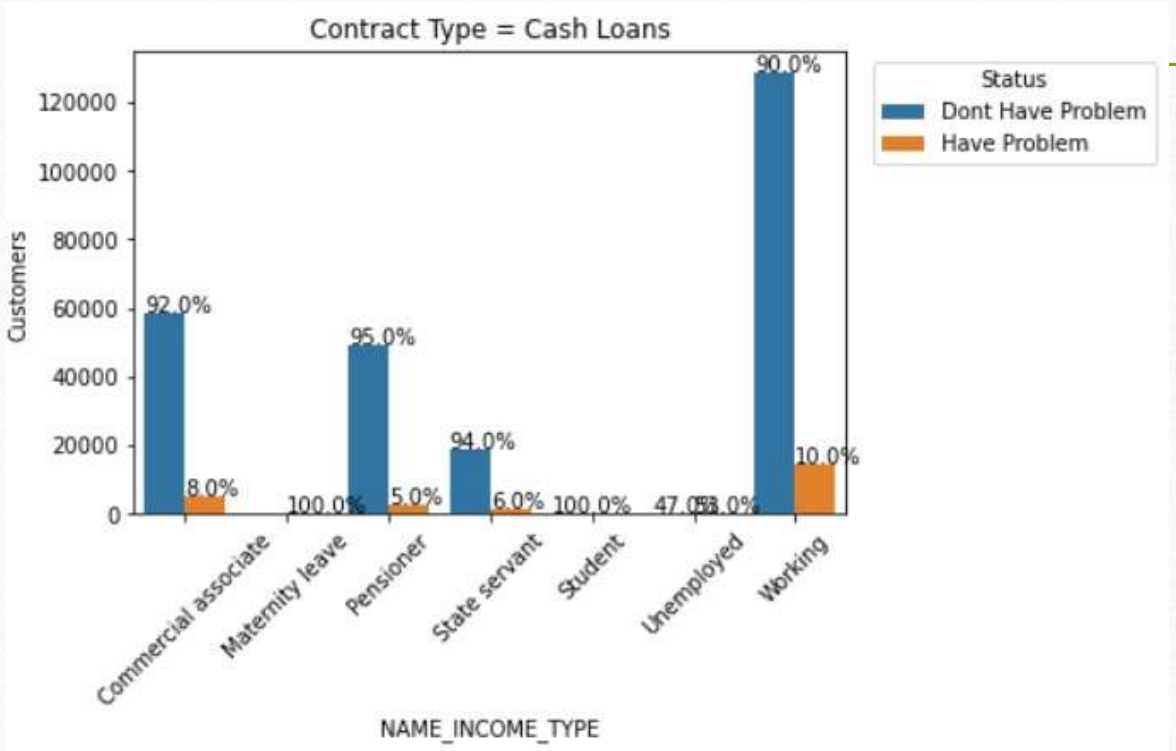
BUSINESS INSIGHT

CONTRACT TYPE (1/2)



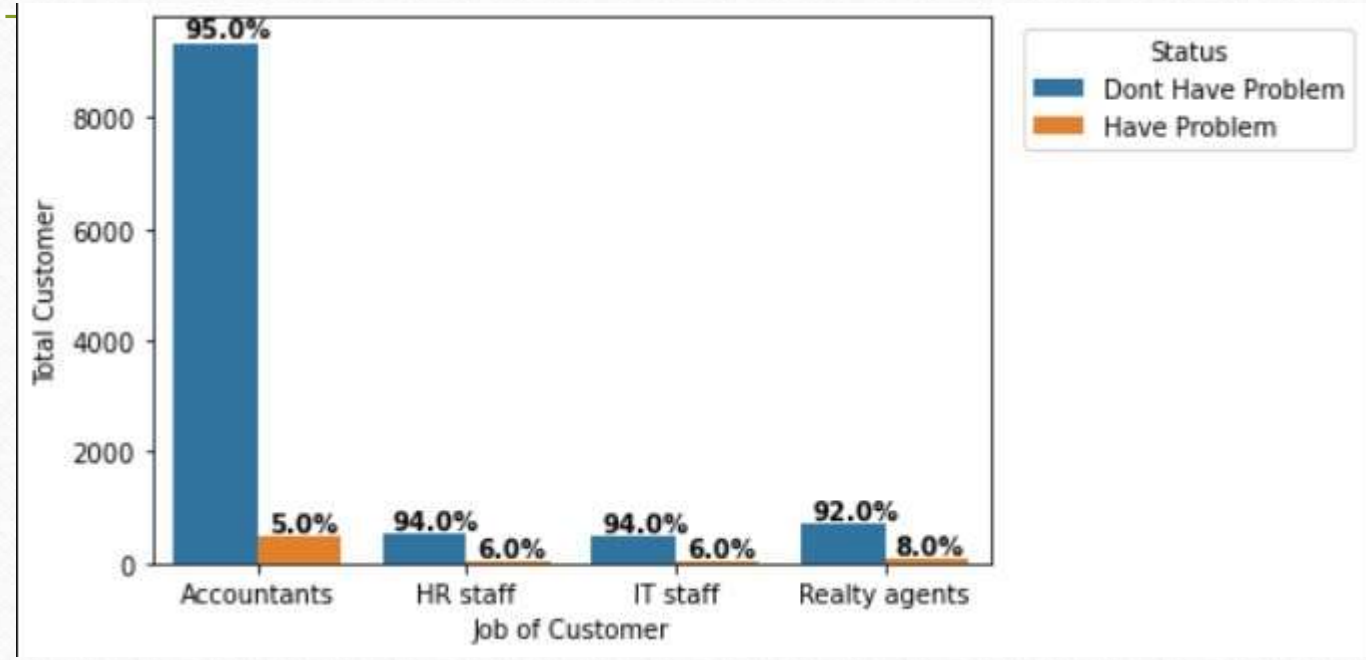
At Revolving loans contract, there are top 4 income type that don't have problem to paying off the loans, namely businessman, maternity leave, student and unemployed, but these 4 incomes have a small number of customers, so we need to increase the number of customers who have income sourced from one of the top 4 income type.

CONTRACT TYPE (2/2)



At cash loans contract, maternity leave is the only one who have 100% customer that have problem to pay off the loan, so we can recommended the maternity leave to try revolving loan because in revolving loan, maternity leave have 100% customer that don't have problem to pay off the loan.

OCCUPATION TYPE OF CUSTOMER



The jobs with the most customers are accountant, so we need to create a campaign as a way of thanking them, then we need to make promotions for the other jobs in graph, because the percentage of customers that have no problem paying the loans is quite high but the number of customer from that three job is still small

MACHINE LEARNING MODEL

STEP BY STEP

Data Preparation

1. Make a new column (age)
2. Feature selection
3. Remove Outlier
4. Feature Encoding (One hot encoding and label encoding)
5. Feature Transformation

Handling imbalance Data

1. Oversampling
2. Undersampling

Train Machine learning model

1. Logistic Regression
2. XGBoost

RESULT

Model	Accuracy Tran Data	Accuracy Test Data	Precision	Recall	F1-score
Logistic Regression with oversampling	94%	94%	94%	94%	94%
Logistic Regression with undersampling	68%	68%	68%	68%	68%
XGBoost with oversampling	96%	95%	96%	95%	95%
Random Forest with oversampling	100%	96%	96%	96%	96%

SUMMARY

1. For Revolving loans contract, we have to find customers with income types of businessmen, maternity leave, students and unemployed.
2. It's recommended to create campaigns for customers who work as accountants because accountants is the largest of number of customers and percentage of successful payments/
3. Create advertisement or promotions for HR staff, IT staff and realty agents to apply for credit.
4. Random forest is the machine learning model chosen to help the team determine whether a customer has a problem paying off a loan or not.