LOAN CREDIT RISK ASSESSMENT

ID/X Partners Internship

By Naufal Rafiawan Basara

TABLE OF CONTENTS

01 **PROBLEM**

Lending company problem overview

02 **DATA PIPELINE**

End-to-end data flow processing

03 **TRAIN & PREDICT**

Selecting the best model to use from their metrics result

04 **CONCLUSION**

Model to use in predicting future value to assess the risk of borrower's loan

05 **OVERVIEW**

What problem has been faced and how to dealt with it

INTRODUCTION

Lending company always faced with various of challenges, especially when it comes to lack of borrowers responsibility. Borrowers comes with different purposes of loan, economic level, and profile.

Therefore, the company have to assess the risk for each of the borrowers to minimize the chance of default to consider which borrower's loan is accepted or not.

Rakamin Academy id/x partners

01

PROBLEM

Lending company problem overview

WHAT IS THE CHALLENGE?

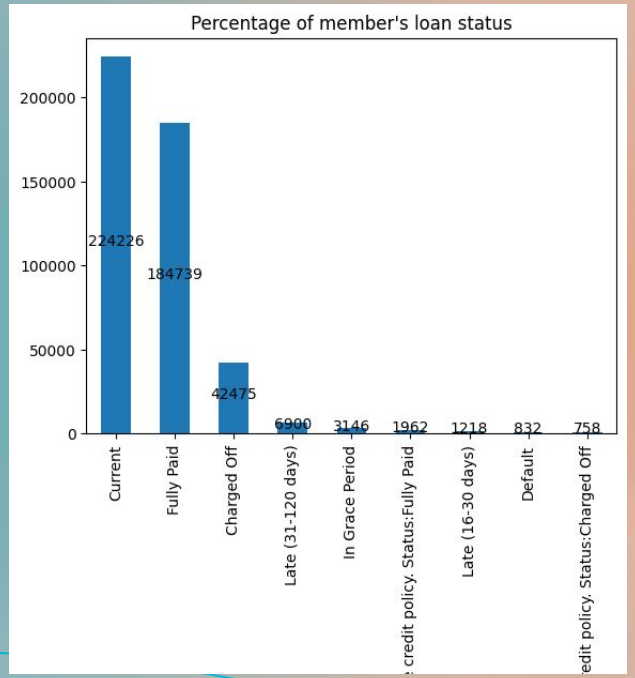
ID/X Partners' client, a credit lending company, is facing a problem to assess the borrower profile. The company have a whole data about borrowers profile in a dataset but still haven't sure about the credit risk assessment, whether the company should accept the loan or not.

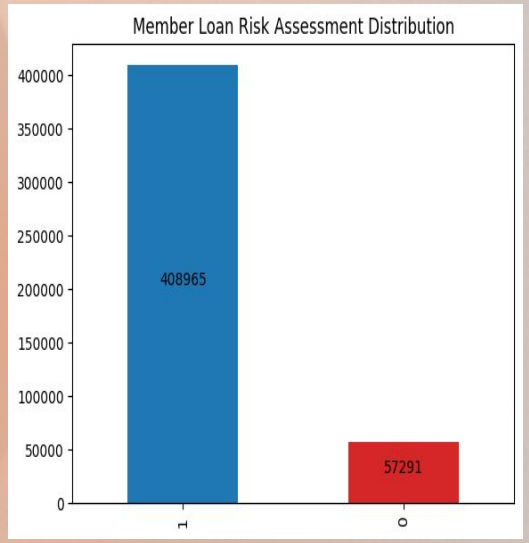
With limited data, the lending company need to predict of what the future borrower's risk possibility is or even the current borrowers risk profile. The dataset given with **74 columns** which include the unique identifier and borrower's profile, current loan status, credit balance, etc. Filled with **466284 data records**, but not all columns has complete data records so that we have to fill it based on information from another column in a dataset if necessary.

HOW IS THE BORROWER' LOAN STATUS?

How do we determine which borrower profile will be accepted or not? First we look at the unique value of the each borrower's loan status. We can classify each of the status to 2 values that indicate the good and risky credit loan.

Here i determine the status is in the good state while the borrowers is on "Current", "Fully Paid", and "Fully Paid even if the borrower does not meet credit policy". Besides that, "Charged Off", "Late", "In Grace Period", "Default" is classify as the risky loan states.





408,965

Number of borrowers with good risk status of loan

57,291

Number of borrowers with risky status of loan

Rakamin Academy id/x partners

02

DATA PIPELINE

End-to-end data flow processing

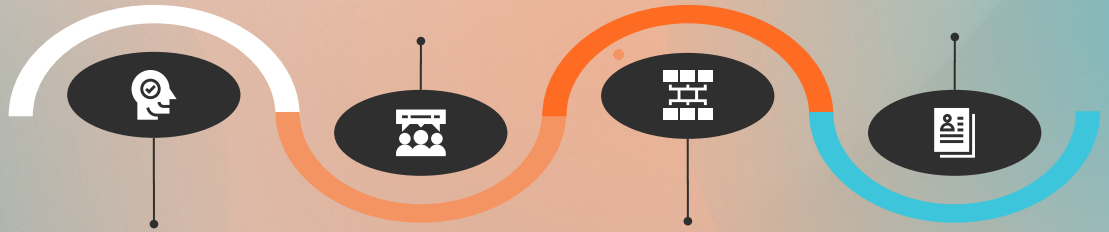
DATA PIPELINE

EDA

Explore important feature to help fitting the model, finding features importance with Random Forest and Decision Tree Regressor

STEP 4

Find and select the optimal model regarding the metrics



PREPROCESS

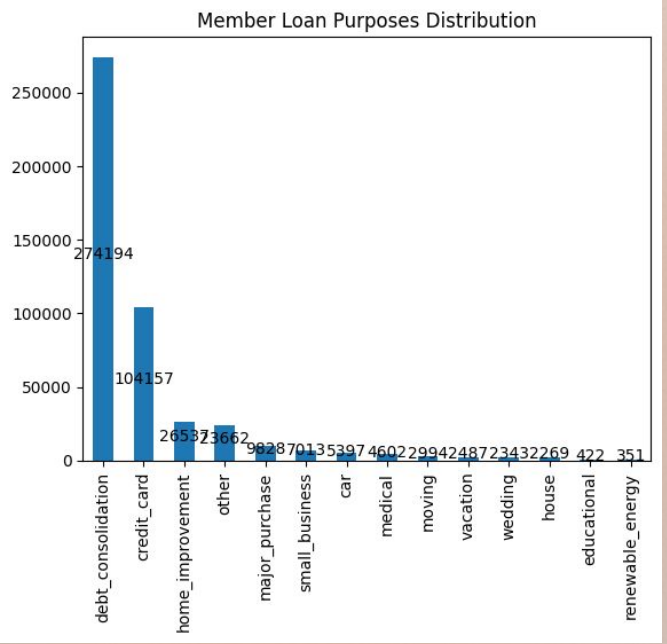
Clean the dataset, such as filling nulls value, remove outliers, Encode the categorical columns, Scaling features

MODELLING

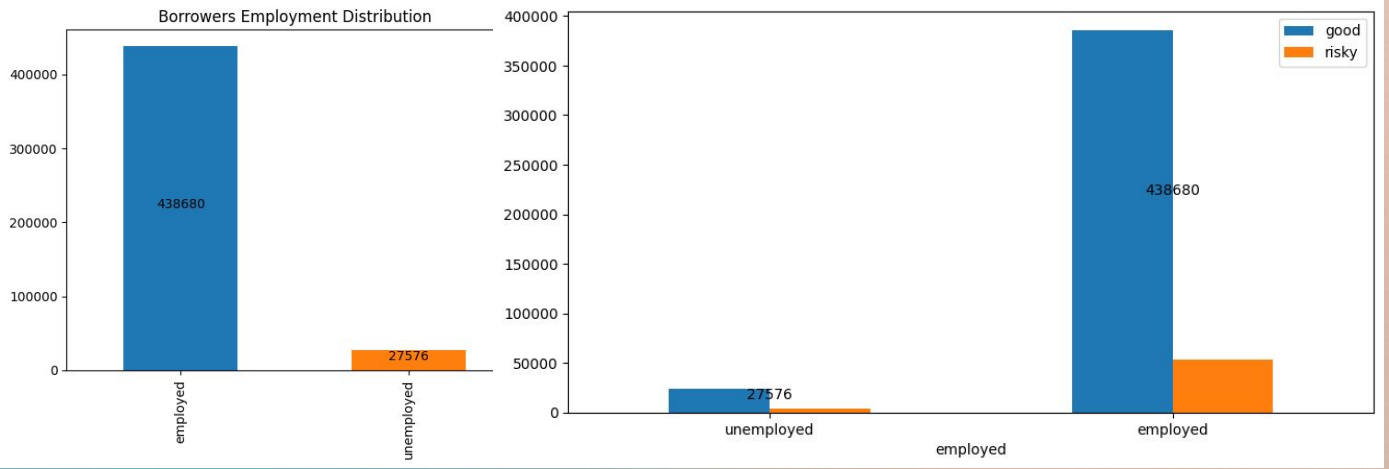
Split the data to two subset train and test set, fit the data to the model.

Borrowers Loan Purposes

The graph states that borrowers tend to borrow in order to consolidate their debt, which leads to risky loan status, follows by credit card with the second highest purposes. Does it mean that the borrowers are not employed since they can't fulfilled the preceding debt?

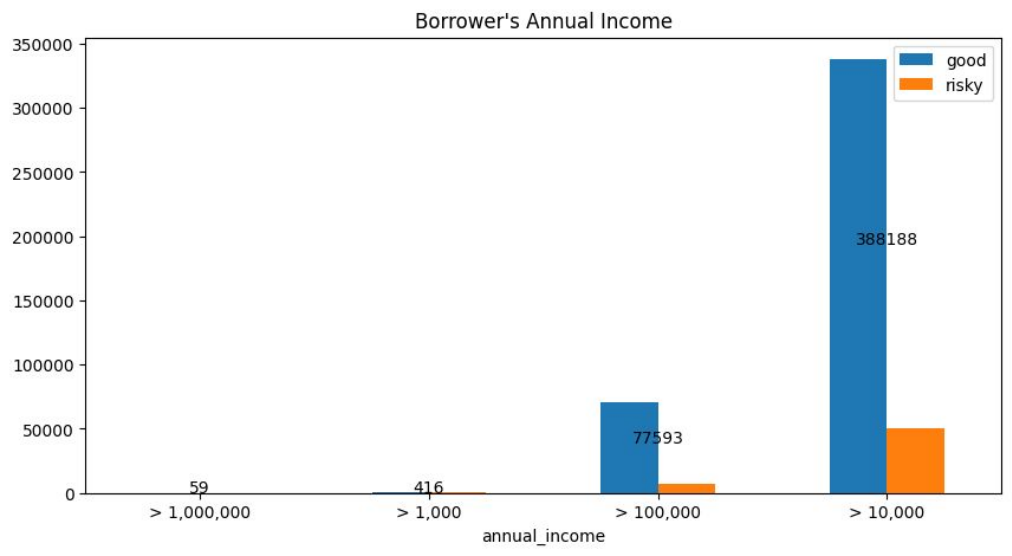


Distribution of the Borrowers Employment Status



Since the borrowers employment status not really showing that unemployed borrower is definitely default, we have to look at each borrower's annual income.

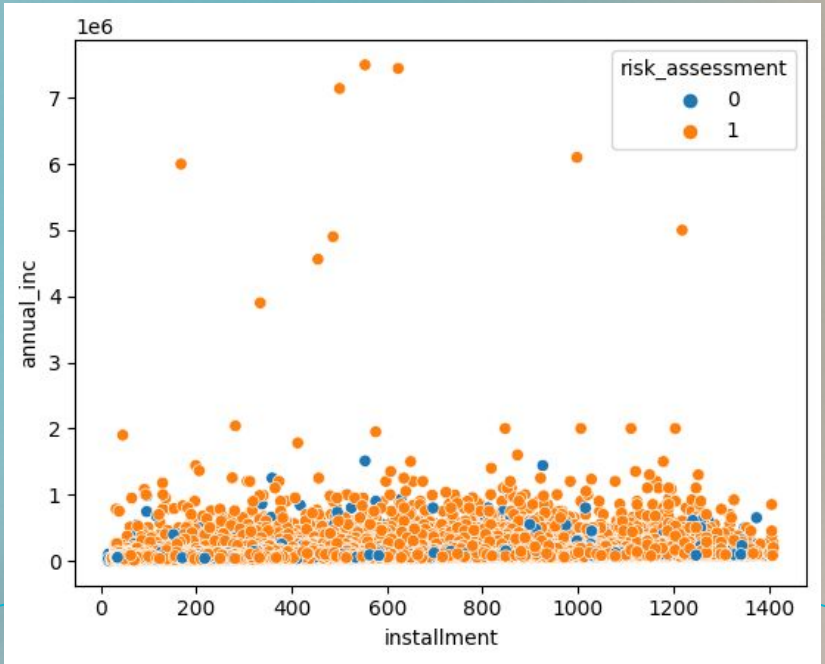
Borrower's Annual Income Distribution



Here, borrowers with annual income higher than 10,000 tends to have more risky loan from the other.

Let's see if annual income affect the installment the borrowers have.

Distribution of Borrower's Annual Income and Installment



But, the installment and annual income that each of the borrower have does not mean if the low annual income and high installment is risky loan.

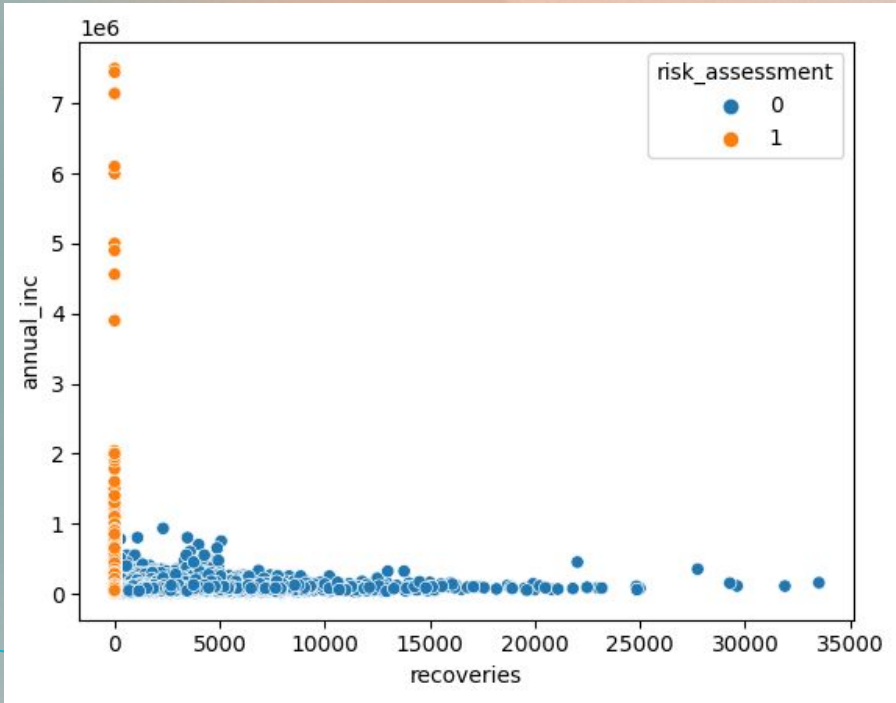
We can't take a conclusion since the distribution are slightly equal.

WHAT IS GOING ON IN THE COMPANY?

But when it comes to comparing borrower's annual income and recoveries, we can see the big difference between the two loan risk status.

The risky loan tends to have more recoveries than good loan.

From the graph, we can conclude that if a **borrower have any recoveries**, the **loan is risky**. It might be happened because of the default, and the lender is taking actions to recover from the loan.



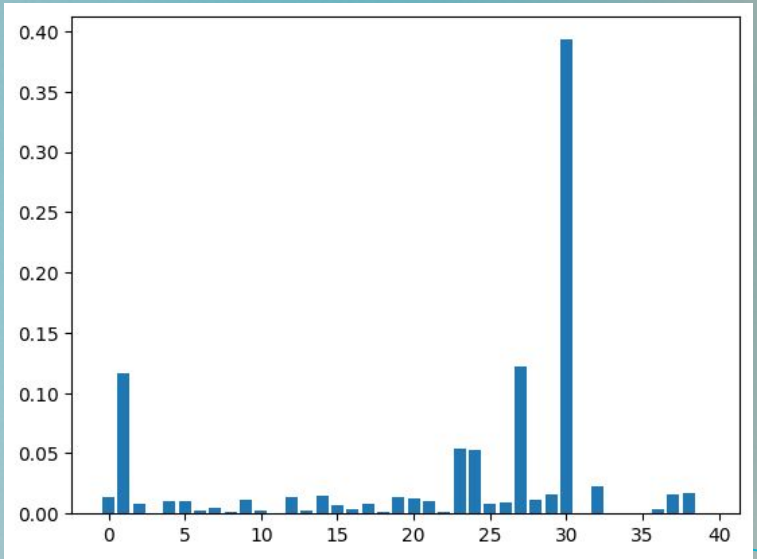
Rakamin Academy id/x partners

03

TRAIN & PREDICT

→ Selecting the best model to use from their metrics result

Feature Importance using Random Forest Regressor



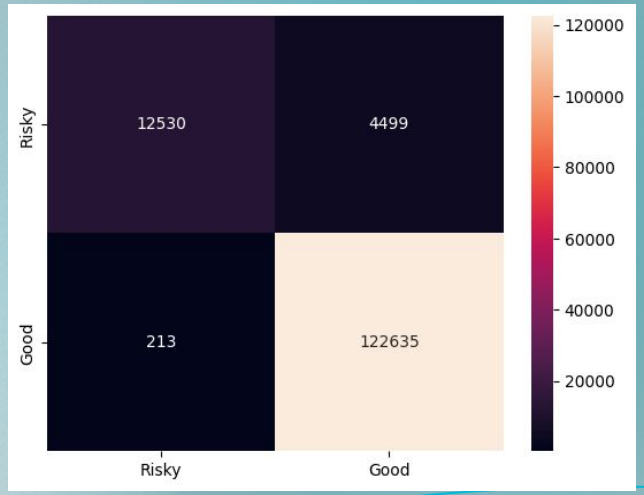
Feature importance calculate score for every feature in a dataset to see whether the feature is relevant or not.

Random Forest regressor catch five features that have slightly significance importance than the others, including **funded_amnt, out_prncp, out_prncp_inv, total_rec_prncp,** and **recoveries**.

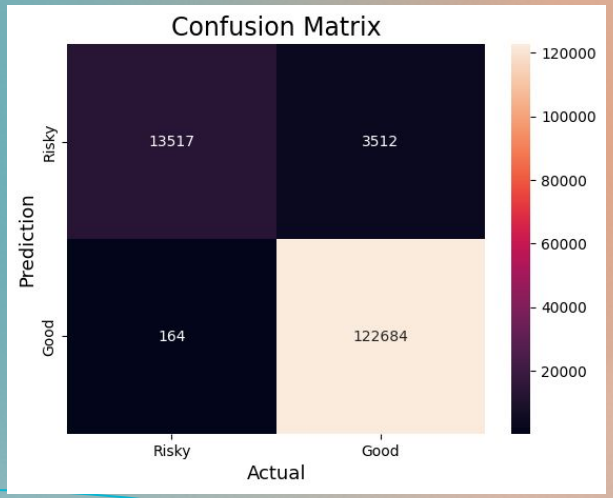
Before the data is being fit to the model, we have to divide the data to 2 sets, which is train and test set. There are a lot of combinations of each size of the subset, but here i used **70% for training** and **30% for testing**.

Trained Models Confusion Metrics

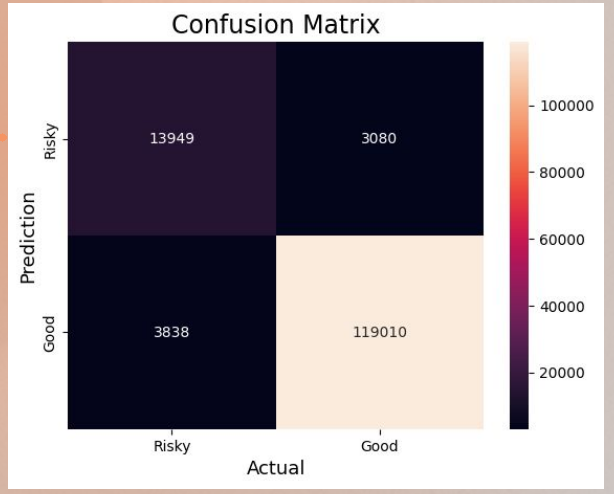
Logistic Regression



Random Forest



Decision Tree



Rakamin Academy id/x partners

04

CONCLUSION

→ Model to use in predicting future value to assess the risk of borrower's loan

MODEL METRICS RESULT

	LOGISTIC REGRESSION	RANDOM FOREST	DECISION TREE
ACCURACY	0.9663	0.9737	0.9505
PRECISION	0.9646	0.9721	0.9747
RECALL	0.9982	0.9986	0.9687
F1 SCORE	0.9811	0.9852	0.9717

Random Forest model is selected due to its metrics result and the highest **accuracy of 0.9737**.

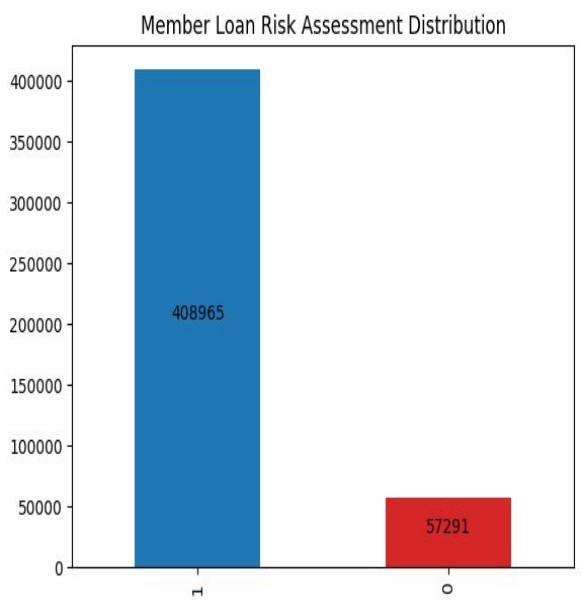
Rakamin Academy id/x partners

05

OVERVIEW

What problem has been faced and how to dealt with it

CONCLUSION



To minimize the chance of borrowers default, lending company need to assess risk status of each of the borrower with the help of machine learning to predict whether the borrower is accepted or not.

Random Forest has been selected as the best model due to its metrics and the highest accuracy of 0.9737 than Logistic Regression and Decision Tree model.

THANKS!

Do you have any questions?

naufalbasara@outlook.com / naufalbasara9@gmail.com
naufalbasara.vercel.app