

# CHAPTER 1

## INTRODUCTION

### 1.1. Background

A loan is a type of debt that is usually in the form of money and is given by individuals, organizations, or financial institutions. The loan is lent to the borrower by adding interest and is repaid on time according to the agreement of both parties. There are two kinds of financial institutions, namely non-banks such as cooperatives, pawnshops, and others, while banks such as central banks, commercial banks, rural credit banks (BPR).

Banks that act as borrowers of money to creditors must have a great risk of experiencing losses caused by creditors who are unable to pay their credit. Therefore, banks must be able to predict which potential creditors can repay their loans and which are unable to provide loans.

To classify creditors who can pay and are unable to pay their debts, this can be seen from several factors such as creditor income, the credit history of prospective creditors, and others. Then these factors are used as material for consideration and analysis to obtain a final decision, whether the credit of the prospective creditor is approved or rejected. We can use this analysis using a computer approach, namely machine learning that can reduce losses experienced by banks.

Therefore, in this project, a classification will be carried out to estimate which prospective creditors are able and unable to pay their credits, with factors such as gender, marital status, education, income, number of credits, a loan term of the prospective creditor. This project implements a logistic regression algorithm, and Extreme Gradient Boosting algorithm.

Logistic Regression is a method used to classify binary data. Logistic regression combines linear features and non-linear or sigmoid functions, which will result in binary data output with a value of 0/1 or completely false. Meanwhile, Extreme Gradient Boosting is an algorithm that works by creating a weak tree so that the new tree focuses on the weakness of the previous tree, which is re-weighted. This model is robust because of the automatic correction after each new tree is added.

In this project, I use machine learning algorithms such as Logistic Regression and Extreme Gradient Boosting to find out which prospective creditors are capable and unable,

applying this computer approach makes it easier for banks to decide whether the results of their credit applications are approved or rejected because the bank can find out which creditors can pay off their debts and which ones don't. The dataset used is taken from Kaggle, where the data is selected, then pre-processed the data that separates the training data and test data, then implements the logistic regression algorithm, Extreme Gradient Boosting on the test data.

## **1.2. Problem Formulation**

1. Can the Extreme Gradient Boosting and Logistics Regression algorithm predict loan approvals?
2. Which algorithm is more suitable for this case, among the Extreme Gradient Boosting and Logistics Regression algorithms?
3. Which variables affect the outcome decision that a loan is rejected and accepted?

## **1.3. Scope**

The data set is taken from Kaggle which contains the results of filling out online application forms from companies that want to automate the loan application process. The dataset has 614 rows and 13 columns containing gender, marital status, education, number of dependents, income, credit, credit history.

## **1.4. Objective**

The purpose of this project is to find out which algorithm is more suitable between Logistic Regression and Extreme Gradient Boosting algorithm for predicting loan approvals and compare which algorithm has the highest accuracy. To know is that the Logistic Regression and Extreme Gradient Boosting algorithm can predict loan approvals. And to know which variables affect the outcome decision that a loan is rejected and accepted.