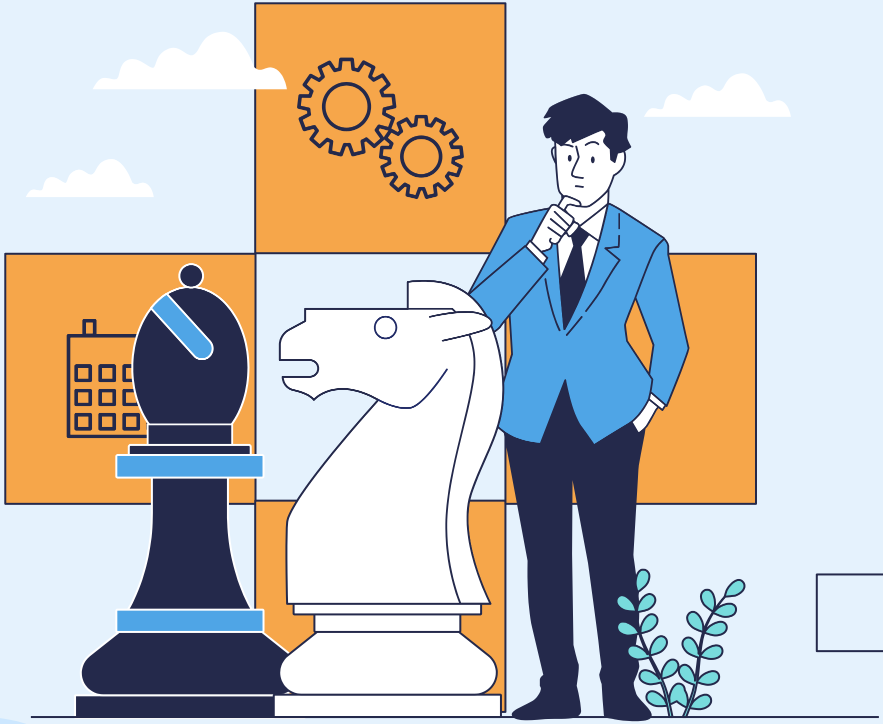


Credit Loan

ID/X Partners

By : Samuel Akwila

Intro

The loan is one of the most important products of the finance's business. All the loan companies are trying to figure out effective business strategies to persuade customers/lenders to apply their loans. However, there are some borrowers behave negatively after their application are approved.

In today's world there are many risks involved in loan companies, so as to reduce their capital loss; companies should perform the risk and assessment analysis of the individual before sanctioning loan. In the absence of this process there are many chances that this loan may turn into bad loan in near future. Loan companies hold huge volumes of lender behavior related data from which they are unable to arrive at a decision point i.e. if an applicant can eligible or not.

For more info:
sakwila96@gmail.com

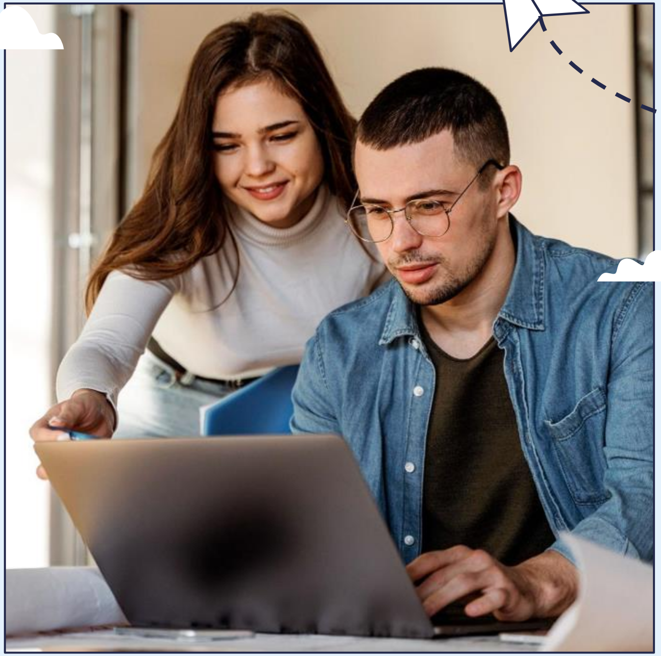
You can visit our notebook:
Link gdrive ipynb

Table of contents

Mission statement

All the keywords you have to know :

- Credit Loan problems
- Binary Classification (Good / Bad Borrower)
- Log Regression, ROC-AUC , Accuracy



Linkedin : linkedin.com/in/samuel-akwila-942a34132
Contact : sakwila96@gmail.com

Samuel Akwila

I am passionate about Data Science and looking for an opportunity to exploit my current skills and become a prominent Data Scientist.

Let’s we talk about this big data

1

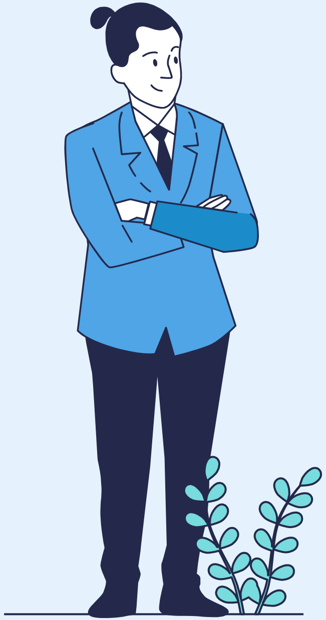
Loan System Overview

Money Loan Company

Problem Statement

Company need to understand customer behavior and predict whether potential customers are eligible for loans and keeping low credit risk as to increase company revenue.

Problem vs solution



Problem

Borrower's bad behaviour after
their application are approved

Solution

Predict the secure loan before
application is approved

2

Data Analysis

Exploratory Data Analysis

Loan Ending : Customer Classification Status

Good borrower ✓	Bad Borrower ✗	Ambiguous ?
Fully Paid	Charge Off	Current
Does not meet the credit policy. Status: Fully paid	Does not meet the credit policy. Status: Charged off	In grace period
-	Default	-
-	Late (16-30 days)	-
-	Late (31-120 days)	-

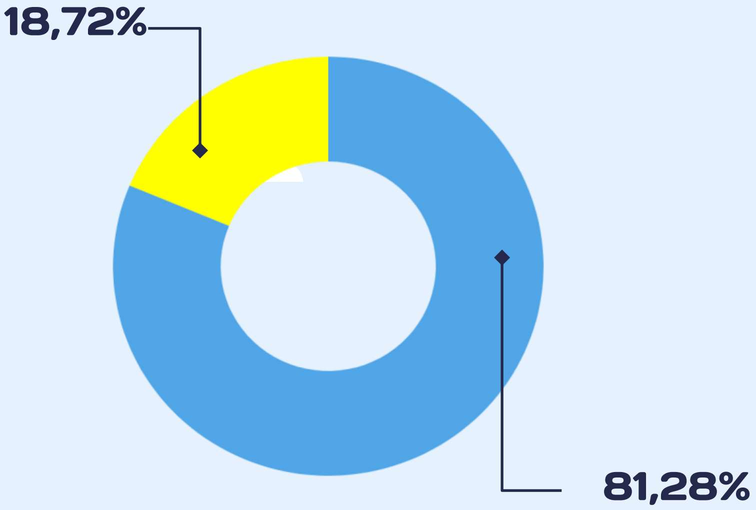
The loan’s problem

18,72%

Bad borrower

81,28%

Good borrower

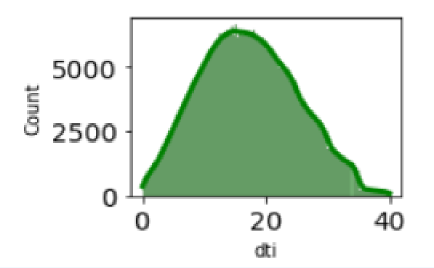


We get the imbalance data which have 81,28% good borrower and 18,72% bad borrower from 2007 until 2014

Exploratory Data Analysis

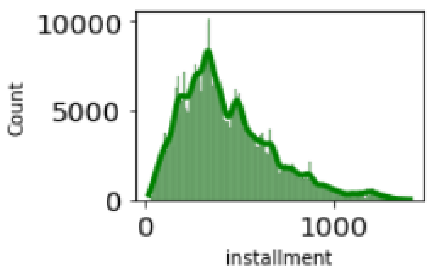
Feature Distribution

Almost all the features contained in the dataset are skewed or have an abnormal distribution. Only the ratio of total obligation (dti) and monthly installments (installments) have normal distribution.



Leakage Data

Some features are indicated as data leaks, because the data was obtained after the loan was in progress. (Issue date, last payment date, and next payment date)



Features observation

75 features

- 20 features **have null values**
- 3 features **have only one value**
- No **duplicate data**
- 5 features need to **change data type** to date
- 28 features **have excess outliers**
- 8 features **have redundant data**
- 15 features **have indicated Leakage data**

Data Preparation

How we handle this data

3

Prepare Data

Handling Missing Value

Fill missing value with median and mode values.

Feature Encoding

Label Encoder

Handling Duplicates

There is no duplicate data

Handling Outliers

Handle 19,07% rows of data

Feature Transformation & Selection

Pick features and transform to right data type

Next Step - Prepare Data

Handling Class Imbalance

SVM SMOTE

Split Data Train & Test

Train : Test = 80 : 20

Machine Learning Model

Decision Tree,
KNN, Gradient Boosting,
Logistic Regression,
Random Forest, XGBoost

Model Prediction

Handling imbalance data

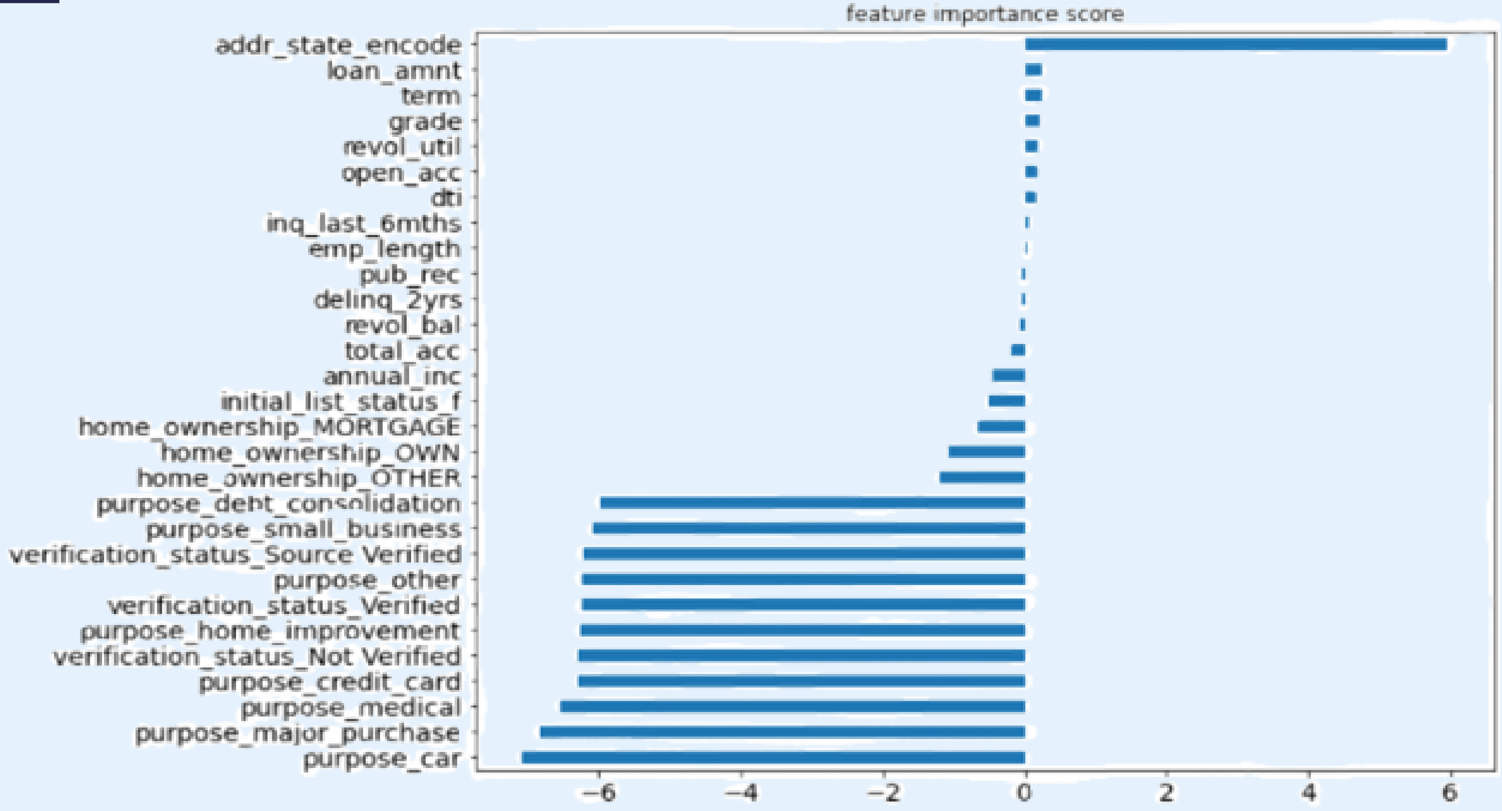
4

Evaluation Models

Model	Akurasi	AUC
KNN	60,86%	57,53%
Decision Tree	70,67%	55,14%
XGBoost	74,10%	59,74%
Gradient Boosting	75,18%	59,43%
Random Forest	78,11%	57,51%
Logistic Regression	70,70%	61,38%
Logistic Regression Tunned	71,00%	61,38%

Predict	Actual	
	Good	Bad
Good	3344	7395
Bad	3801	23636

Feature Importance



addr_state_encode	The state provided by borrower
purpose_debt_consolidation	Combining purpose loans in the form of consolidation
purpose_small_business	Purpose for small business
verification_status_source_ verified	Verification of source income status
loan_amnt	Amount of money borrowed
verification_status_verified	Verification borrower status
purpose_home_improvement	Purpose loans for upgrading home
verification_status_not_verified	Verification status that’s not verified
purpose_credit_card	Purpose loans for credit card
purpose_medical	Purpose loans for medication
purpose_major_purchase	Purpose loans for individual needs
purpose_car	Purpose loans for car

5

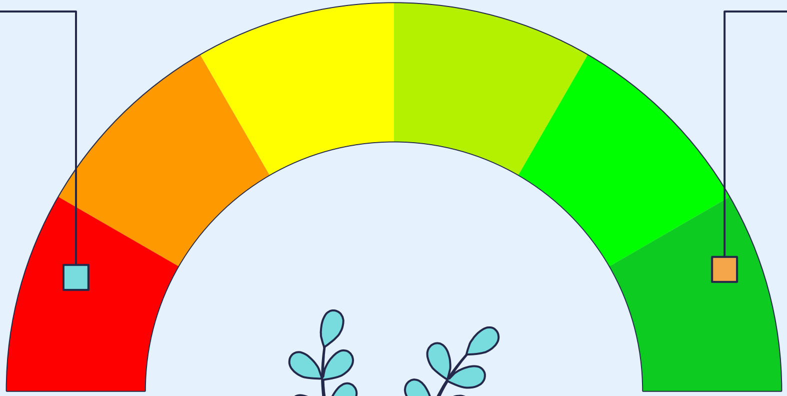
Solution

Final product of this project

Predict the Borrower With Credit Score

Bad Borrower

Probability of bad borrower
< 0.500



Good Borrower

Probability of good borrower
> 0.500

Before - After : Bad Loan

21,8 %

6,3 %

Saving money from the Bad Borrower

Bad borrower rate potentially decrease by -71.0% from the previous 21.8% to 6.3% after action based on predictive modeling as 37052 borrower (before model there were 52168 bad borrower)

6

Action Item

Recommendations

Action Items

Hope this project could help this company to improve :

Feature Importance	Characteristics (value)	Action Item
Verified status source	1	Prioritized borrower that has verified status source.
Purpose	1	Prioritized borrower who purpose for debt consolidation, credit card, small business, buy new car, medical tuition, individual need, and home improvement.
Address state	-	Prioritized borrower who lives or settle in good real estate or residence.
Loan amount	-	Review amount that borrowed with the purpose, income salary, and address state.

Additional Data for Further Research

experienced	**birth_date**	**dependence**
To know if the borrower have another loan in other company or in the past	To know customer's age	To know about borrower's dependences
marital_status	**expenses**	**...**
To know borrower's marital status	To know borrower's expenses every month	

Thanks!