CREDIT RISK PREDICTION

By Aditya Bagus Pratama

[Notebook Link]

PRELIMINARY

- In every company, especially **Lending Company**, a **smooth transaction and credit**, especially in paying off a loan, is **extremely important**
- **Bad credit**, mainly due to defaulted borrowers, is **the biggest loss factor** for Lending Companies
- Losses due to borrowers who fail to pay, are actually caused by the **selection** of prospective borrowers who are **not quite right**

GOAL

A Lending Company **needs to have a selection system** that can **predict and classify** prospective borrowers who are safe or at risk, so that lending can be **right on target.**

OBJECTIVES

We will utilize **machine learning** technology that works quickly in knowing the patterns and characteristics of borrowers, to **determine** whether a prospective borrower is **eligible for a loan** or not.

BUSINESS METRICS

- **Net Profit Margin**

  company finances will be more stable
- **Return of Assets**

  targeted lending, making the capital issued more effective
- **Minimum Loss**

  Losses will be reduced due to targeted loans

DATA UNDERSTANDING

- The data that will be used for modeling is the **Loan Dataset from 2007-2014**
- There is "target" features which contains the categorization of the borrower based on their **“Loan Status”**, if the target value is "1" then the borrower has successfully paid, if "0" the borrower is late or failed to pay.

THE BORROWER'S CHARACTERISTICS

- **Loan Status (loan_status)**

  Loan Status is feature that will be used to make "Target" features which contains status in which loan is categorized as bad or good and should be accepted or not

- **Purpose (purpose)**

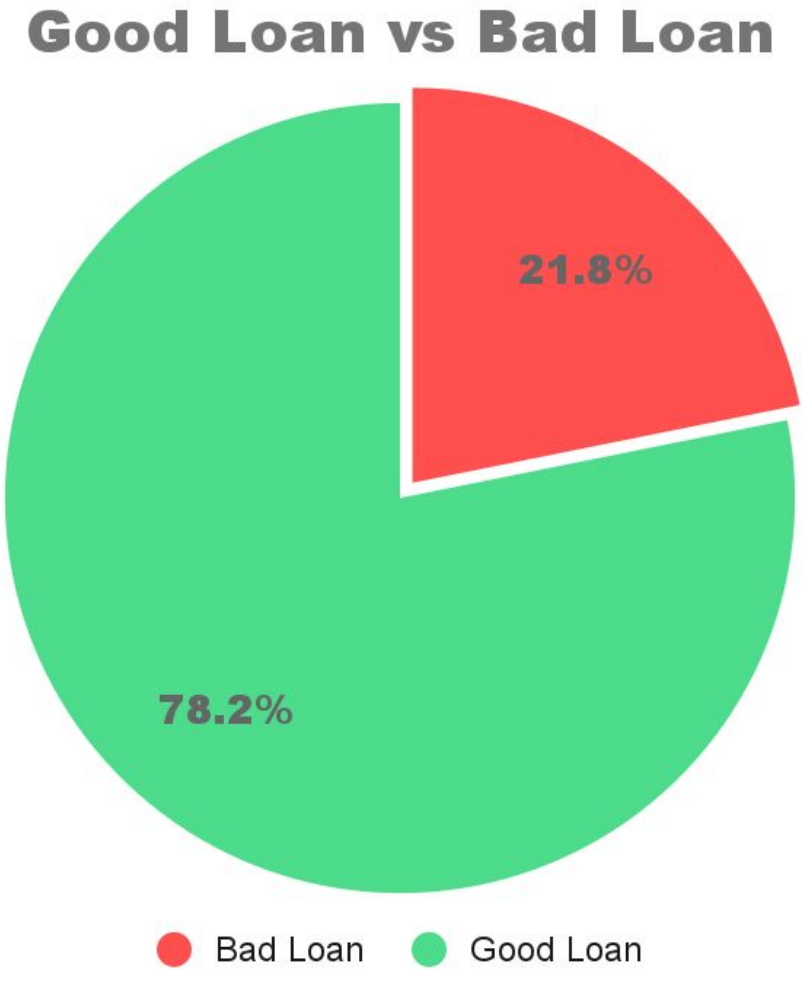
  Purpose is contains the purpose of borrowers in applying loan/credit

- **Loan Amount (loan_amnt)**

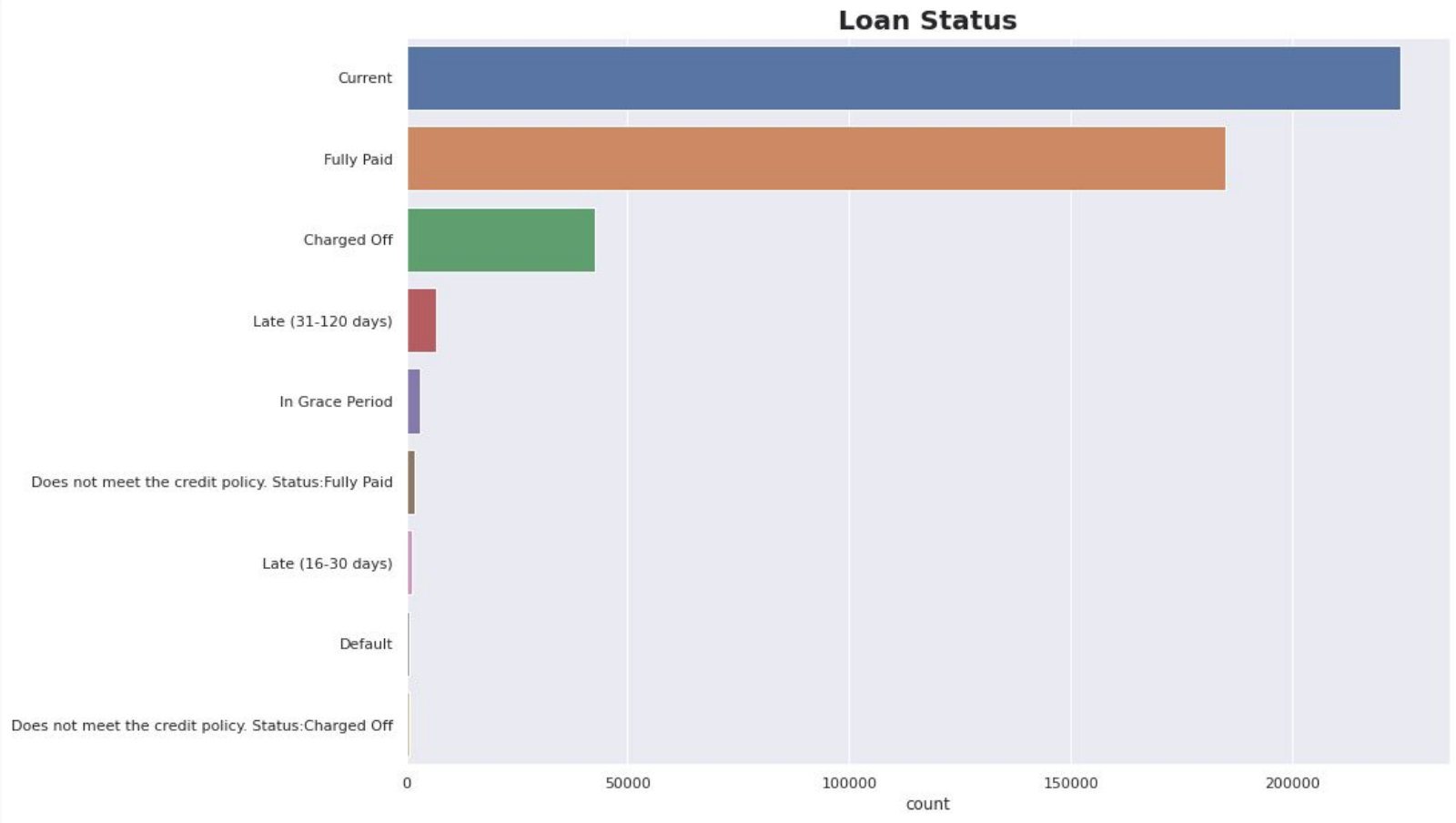
  It contains the amount of loan applied by the borrowers

- **Loan Grade (grade)**

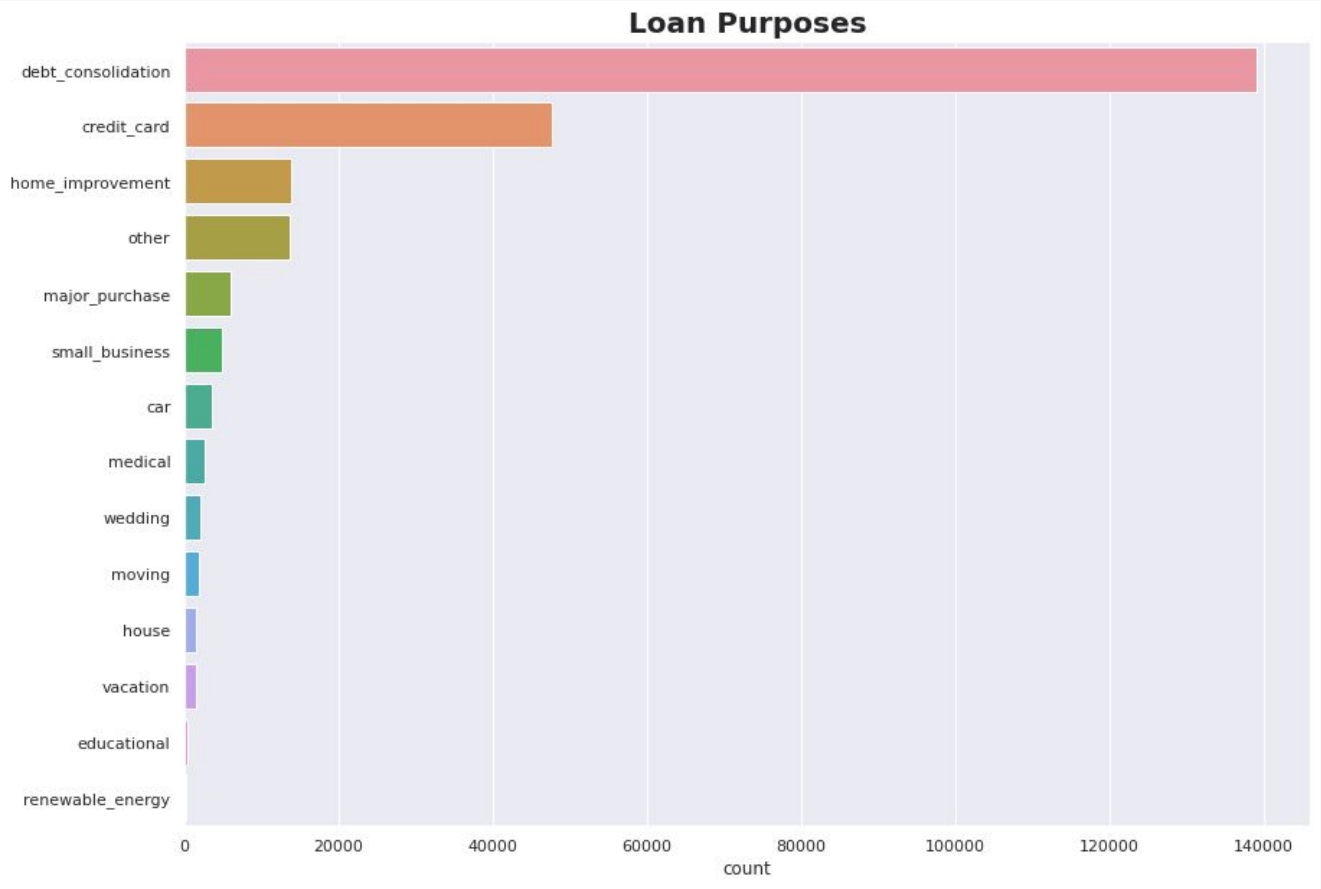
  It ranks the loan alphabetically from **A to G**, in which the bigger a loan close to G grade, the bigger interest will be get


Good Loan vs Bad Loan
21.8%
78.2%
Bad Loan
Good Loan

Loan Status View



Loan Purpose View



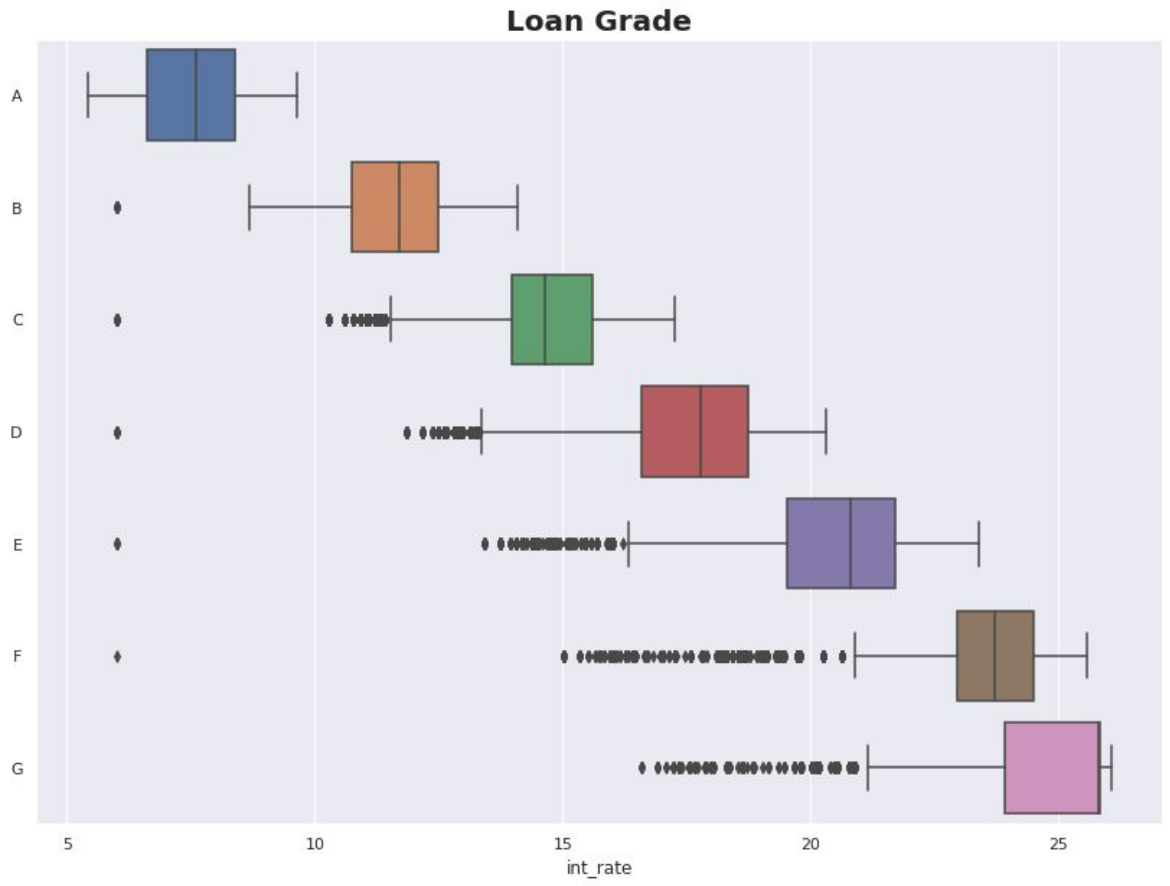
The majority purpose of loan is for Debt Consolidation

Loan Amount Distribution View

Loan Status	Mean	Std
Good Loan	13,167	7,944
Bad Loan	14,624	8,390

The Mean and Standard Deviation value in Bad Loan is bigger than in the Good Loan

Loan Grade View



Each grade has its own interest rate

Grade A with **7.5%** Interest Rate up to **Grade G** with **24%** Interest Rate

DATA PREPARATION

Some preparations have done to before modeling process, are:

- Dropping some columns with total and >40% missing values.
- Filling up some columns with <40% missing values by assigning mode into categorical data and median into numerical data.
- Removing duplicated data
- Removing some columns with less correlation into Loan Status
- Doing some feature engineering for columns data type

MODELING

The data are splitted into **Train Data** and Test Data with **80:20** comparison.

The model uses **Logistic Regression** since we need to do classification with binary target

(good or bad)

ACCURACY RESULT

The model can predict loan status precisely with >95% accuracy

96.1%

for Train Data

95.9%

for Test Data

MODEL PERFORMANCE

PRECISION

87%

Possibility performance to predict a loan is a bad loan in prediction, and is actually bad loan in reality

RECALL

96%

Possibility performance to not wrong in predicting a bad loan as a good loan

F1-SCORE

87%

Combination of Recall and Precision to state the consistency of model in predicting loan status

MODEL ANALYSIS

By using **Logistic Regression** algorithm, the model has reach **96%** accuracy in predicting loan status. It means, the model can analyze the borrower's characteristics and classify them into loan status **very well.** So basically we can use this model in the business process, so that our **business metrics can be fulfilled,** which means the business risk and loss can be minimized, even profit can be increased.

CONCLUSION

BUSINESS BENEFITS

The model can be used to fulfilled business metrics

BORROWER'S CHARACTERISTIC

Borrowers with bad loan status tend to borrow larger loan amounts than customers with good loan status

PREDICTION RESULT

Model can be used to predict and determine the loan status really well

adityabagusp345@gmail.com

github.com/bytadit

linkedin.com/in/aditya-bagus-pratama

THANKS !