LOAN CREDIT RISK PREDICTION

Presented by
Gilang Wiradhyaksa

TABLE OF CONTENTS

01

ABOUT THE PROJECT

A predictive machine learning project

PROJECT INTRODUCTION

In the dynamic landscape of financial services, lending institutions face the formidable challenge of balancing risk and reward when extending credit to borrowers.

Our project focuses on leveraging machine learning algorithms to predict the credit risk associated with loan applicants.

PROBLEM

Manual prediction relies on human judgment, which can be subjective and prone to bias.

Manual prediction is inherently limited by human capacity and resources.

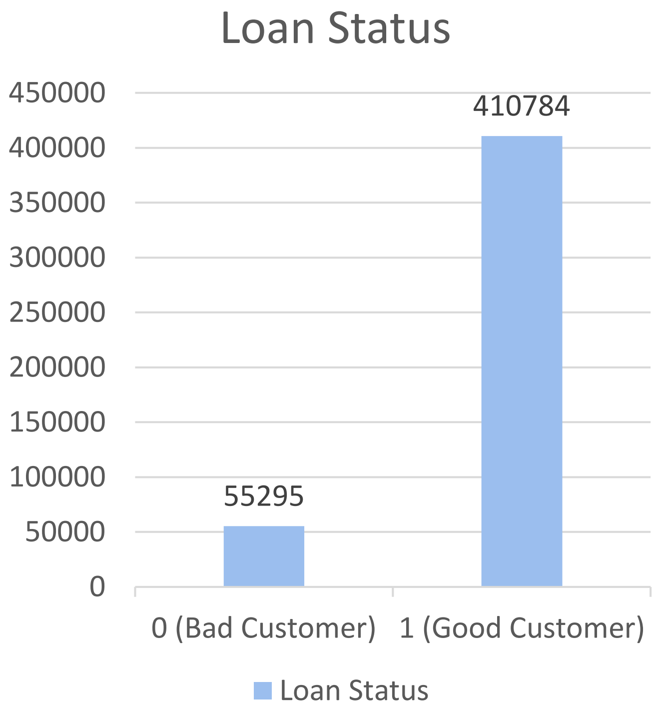
Manual prediction lacks consistency across different decision-makers and time periods.

GOALS

In contrast with manual prediction, machine learning approaches offer several advantages, including scalability, objectivity, efficiency, and adaptability.

Machine learning models can analyze vast amounts of data, identify complex patterns, and provide more accurate and consistent predictions of loan credit risk.

LOAN STATUS DISTRIBUTION



- The data is imbalance, most of the target are positive.
- Imbalance data is normal for credit case, because normally most of customer always paying their debt

MODEL EVALUATION

Three Best Algorithm

Logistic Regression

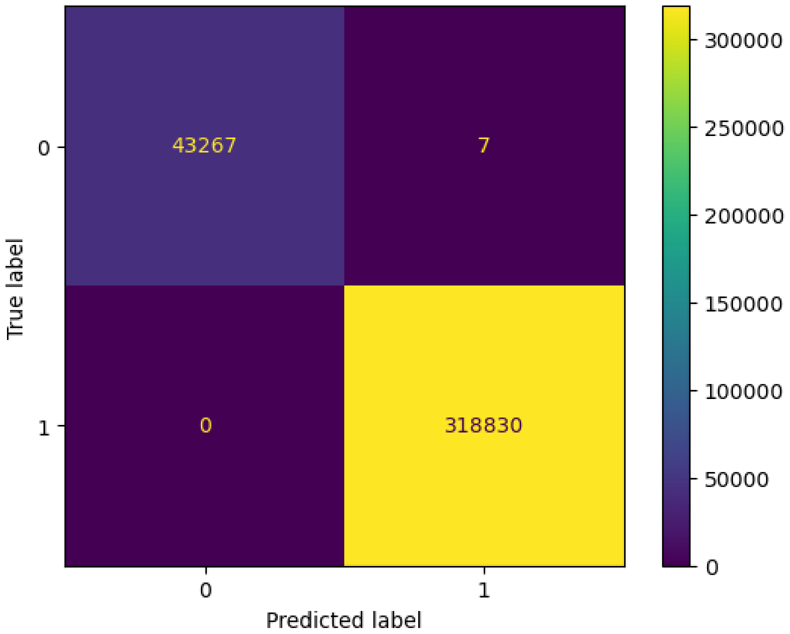
Precision Score : 90%

K-Nearest Neighbors

Precision Score : 93%

Random Forest

Precision Score : 99%



Random Forest Confusion Matrix

You can check this project here!

MODEL ANALYSIS

Metrics

To analyze this model, we use **Precision** metrics because we want to **minimalize False Positive**.

In this case False Positive is customer whose **actually Bad** or Negative but **predicted as Good** Customer or Positive.

False Positive

Why is it better to **minimalize** False Positive ?

For this loan case, it is **better** if the **Lending Company reject good customer** rather than **provide loan to a bad customer** who have no intention or unable to pay.

Cross Validation

After performing Cross Validation check on **Random Forest** model we still get very good result for the model which mean it's **not a coincidence** this model have a very good result which is 99% Precision Score.

You can check this project here!

CONCLUSION

This dataset is having **imbalance data**. For a better result this needs to be handled. Technique like **SMOTE** can be used to handle imbalance dataset.

Out of 3 model we tried, the best model is **Random Forest**. This model give 99% Precision score which mean it can correctly predict 99 out of 100 data.

This model weakness is even with 99% precision score, when we check using **cross validation** we only get about **94% of precision score**. Less than the train result but it's still very good result.

For next **improvement**, we can try **oversampling** data using **SMOTE** technique and try to do **hyperparameter tuning** for our model. With those we might be able to improve the model even better.

You can check this project here!

ABOUT ME

Gilang Wiradhyaksa

Data Scientist

I am a Fullstack Software Developer transitioning to Data Science. Experienced in building a Web Application, API Web Service and Background Service.

Possess an understanding of statistical analysis, machine learning, and data visualization techniques, combined with strong programming skills in Python and proficiency in SQL.

My experience includes data preprocessing, feature engineering and model development.

LinkedIn : in/gilangwiradhyaksa/

Email : gilang.wirad@gmail.com

GitHub : github.com/gilangwd

THANK YOU