Building a Model for Predicting Credit Risk

Dwi Putra Satria Utama

Introduction

Challenges in Managing Credit Risk, Especially in Regards to Loan Default/Failure in Payments.

Objectives

The objective is to develop a machine learning model that can accurately predict credit risk for debtors.

Why is it important?

Because credit risk has significant financial implications for companies.

Because with the right recommendations, the company can take necessary actions to reduce the risk of loan defaults.

Dataset

Rakamin-ID/X Partners Credit Risk Dataset :
https://rakamin-lms.s3.ap-southeast-1.amazonaws.com/vix-assets/idx-partners/loan_data_2007_2014.csv

466285 rows 75 columns

Filter Data for Credit Risk Prediction Model

96502 rows 15 columns

The credit risk prediction model focuses on early prevention by selecting key features during loan applications. Filters are applied to identify low-risk borrowers and manage risk effectively. Feature selection includesterm_months', 'grade', 'sub_grade', 'emp_length', 'emp_title', 'home_ownership', 'verification_status', 'purpose', 'addr_state', 'initial_list_status', 'loan_status' with 'loan_category' as the target. Missing values are handled by deletion due to MCAR, and undersampling is applied for target balance. Outliers are managed using IQR Winsorization to maintain the data distribution essential for logistic regression stability.

Best Model

0 = GOOD CATEGORY

1 = BAD CATEGORY

Train set

Logistic Regression		
Kelas	Recall	Accuracy
0	1.00	1.00
1	1.00	

Decision Tree		
Kelas	Recall	Accuracy
0	1.00	1.00
1	1.00	

Neural Network with Keras (Sequential)		
Kelas	Recall	Accuracy
0	1.00	1.00
1	1.00	

Test set

Logistic Regression		
Kelas	Recall	Accuracy
0	1.00	1.00
1	1.00	

Decision Tree		
Kelas	Recall	Accuracy
0	1.00	1.00
1	1.00	

Neural Network with Keras (Sequential)		
Kelas	Recall	Accuracy
0	1.00	1.00
1	1.00	

Best Model

The logistic regression model, decision tree, and neural network with keras (Sequential) exhibit equally good performance. However, the decision tree model outperforms in terms of computational efficiency.

Consequently, I have chosen the decision tree as the best-performing model for saving and later use in the deployment phase.

Conclusion

The Decision Tree model is the best model for predicting credit risk .

Business Implementation

1. Its implementation improves the efficiency of debtor assessments.
2. It functions as an early warning system to detect potential high-risk debtor.
3. It assists in making accurate decisions based on data analysis, reducing the risk of errors, and minimizing financial and business reputational impacts.
4. It allows for the review of loan requirements, offers alternative solutions for high-risk customers, and allocates more resources for credit monitoring and debt recovery.

Deployment

Link Deployment