

ABSTRAK

Dalam industri perbankan, risiko gagal bayar pinjaman (*Non-Performing Loan/NPL*) menjadi tantangan besar yang dapat mempengaruhi stabilitas keuangan suatu lembaga. Penelitian ini bertujuan untuk mengembangkan model klasifikasi risiko gagal bayar menggunakan algoritma *Light Gradient Boosting Machine* (LGBM), yang dikenal memiliki keunggulan dalam kecepatan pemrosesan dan akurasi tinggi, terutama dalam menangani data tidak seimbang.

Penelitian ini menggunakan dataset publik dengan 255.347 data nasabah dan 18 atribut. Model dikembangkan menggunakan tiga pendekatan: tanpa *balancing*, *balancing* dengan *Synthetic Minority Oversampling Technique* (SMOTE), dan *balancing* dengan parameter LGBM, serta dilakukan *hyperparameter tuning* untuk mengoptimalkan performa. Hasil eksperimen menunjukkan bahwa model tanpa *balancing* memberikan akurasi tertinggi sebesar 88,75%, diikuti oleh *balancing* dengan SMOTE (85,09%), sementara *balancing* menggunakan parameter LGBM memiliki akurasi lebih rendah sebesar 70,56%. Dari segi efisiensi, model tanpa *balancing* juga memiliki waktu komputasi lebih cepat (39,0382 detik), dibandingkan dengan *balancing* menggunakan SMOTE (60.8197 detik) dan LGBM dengan parameter *balancing* (36.4757 detik).

Dengan hasil ini, model LGBM tanpa *balancing* terbukti menjadi solusi terbaik untuk mendeteksi risiko gagal bayar dengan akurasi tinggi dan waktu pemrosesan yang optimal. Penerapan model ini dapat membantu perbankan dalam meminimalkan tingkat NPL, meningkatkan strategi mitigasi risiko, serta mendukung pengambilan keputusan kredit yang lebih akurat.

Kata Kunci: *Light Gradient Boosting Machine*, *Non-Performing Loan*, Klasifikasi Kredit, *ensemble Learning*, Risiko Kredit.

ABSTRACT

In the banking industry, the risk of loan default (Non-Performing Loan/NPL) poses a significant challenge that can impact the financial stability of an institution. This study aims to develop a loan default risk classification model using the Light Gradient Boosting Machine (LGBM) algorithm, which is known for its high processing speed and accuracy, especially in handling imbalanced data.

This research utilizes a public dataset containing 255,347 customer records with 18 attributes. The model is developed using three approaches: without balancing, balancing with the Synthetic Minority Oversampling Technique (SMOTE), and balancing using LGBM parameters. Additionally, hyperparameter tuning is performed to optimize performance. Experimental results show that the model without balancing achieves the highest accuracy of 88.75%, followed by balancing with SMOTE (85.09%), while balancing using LGBM parameters yields a lower accuracy of 70.56%. In terms of efficiency, the model without balancing also has the fastest computation time (39,0382 seconds) compared to balancing with SMOTE (60.8197 seconds) and LGBM with balancing parameters (36.4757 seconds).

Based on these findings, the LGBM model without balancing proves to be the best solution for detecting loan default risk with high accuracy and optimal processing time. The implementation of this model can help banks minimize NPL rates, enhance risk mitigation strategies, and support more accurate credit decision-making.

Keywords: *Light Gradient Boosting Machine, Non-Performing Loan, Credit Classification, Ensemble Learning, Credit Risk.*