ABSTRAK

Stacyana Jesika Surianto, NIM 4211250002 (2021), Perbandingan Kinerja Model Logistic Regression Dan XGBoost Dalam Memprediksi Skor Kredit (*Credit Scoring*)

Credit Scoring adalah pendekatan matematis yang digunakan untuk menilai kelayakan kredit individu atau perusahaan dengan mengklasifikasikan debitur ke dalam kategori tertentu berdasarkan profil risiko debitur. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *machine learning* yaitu Logistic Regression dan XGBoost dalam memprediksi skor kredit (*credit scoring*) untuk mengurangi risiko *Non-Performing Loan* (NPL) pada PT Graha Mazindo Mandiri. Dataset sekunder yang digunakan berisi 1.533 data debitur kredit mobil dengan 17 variabel, diantaranya 1 variabel terikat dan 16 variabel bebas. Proses penelitian meliputi *preprocessing data* (pembersihan, penanganan *outlier*, *encoding*, normalisasi, dan penyeimbangan kelas dengan SMOTE), pemodelan, dan evaluasi menggunakan metrik *Accuracy, Precision, Recall, F1-score*, dan ROC-AUC. Hasil menunjukkan XGBoost unggul dengan akurasi 98% dan ROC-AUC 0,99 dibandingkan Logistic Regression dengan akurasi 90% dan ROC-AUC 0,96, karena kemampuan XGBoost menangkap pola non-linier dan menangani ketidakseimbangan data. Penelitian ini memberikan wawasan tentang faktor risiko kredit dan mendukung pengambilan keputusan kredit yang lebih akurat, dengan rekomendasi untuk optimalisasi *hyperparameter* dan integrasi model ke dalam sistem operasional.

Kata Kunci: *Credit Scoring*, Logistic Regression, XGBoost, *Machine Learning*, *Non-Performing Loan*, SMOTE

ABSTRACT

Stacyana Jesika Surianto, NIM 4211250002 (2021), Comparison of Logistic Regression and XGBoost Model Performance in Predicting Credit Scores

Credit Scoring is a mathematical approach used to assess the credit worthiness of individuals or companies by classifying debtors into specific categories based on their risk profiles. This study aims to compare the performance of Logistic Regression and XGBoost machine learning algorithms in predicting credit scores to reduce the risk of Non-Performing Loan (NPL) risk at PT Graha Mazindo Mandiri. The secondary dataset used contains 1,533 car loan debtor data with 17 variables, including 1dependent variable and 16 independent variables. The research process includes data preprocessing (cleaning, handling outliers, encoding, normalization, and class balancing with SMOTE),modeling, and evaluation using the Accuracy, Precision, Recall,F1-score, and ROC-AUC metrics. The results show that XGBoost excels with 98% accuracy and ROC-AUC of 0.99 compared to Logistic Regression with an accuracy of 90% and ROC-AUC 0.96, due to XGBoost's ability to capture non-linear patterns and handle data imbalance. This study provides insights into credit risk factors and supports more accurate credit decision-making, with recommendations for hyperparameter optimization and model integration into operational systems.

Keywords: Credit Assessment, Logistic Regression, XGBoost, Machine Learning, Non-Performing Loans, SMOTE