ABSTRAK

Pertumbuhan layanan pinjaman *online* di Indonesia membawa tantangan signifikan berupa risiko keterlambatan pembayaran (*late payment*) dan gagal bayar (*default*). Penelitian ini bertujuan membandingkan performa algoritma *Random Forest* dan *XGBoos*t dalam memprediksi status pembayaran nasabah menggunakan *dataset* historis dari *Lending Club*. Tantangan utama dalam data ini adalah ketidakseimbangan kelas (*imbalanced data*), di mana nasabah tepat waktu jauh lebih banyak dibandingkan nasabah *default*. Metodologi penelitian meliputi tahap praproses data seperti pembersihan, penanganan nilai hilang, normalisasi menggunakan *StandardScaler,* serta penanganan ketidakseimbangan kelas melalui teknik SMOTE dan *Class Weight*. Model dioptimasi menggunakan *GridSearchCV* dan *Threshold* Tuning untuk mendapatkan ambang batas keputusan yang ideal. Hasil evaluasi menunjukkan bahwa skenario *XGBoost + GridSearchCV + Threshold* Tuning merupakan model paling optimal dengan nilai akurasi mencapai 78,87%, presisi 58,98%, dan nilai ROC-AUC sebesar 0,7211. Meskipun *Random Forest* dengan *Class Weight* memiliki *recall* lebih tinggi, *XGBoost* lebih unggul dalam menjaga ketepatan prediksi dan menekan angka *False Positive*. Variabel tingkat suku bunga (int_rate), *grade,* dan rasio beban utang (dti) ditemukan sebagai indikator risiko paling signifikan. Implementasi sistem melalui prototipe berbasis *Streamlit* membuktikan bahwa integrasi aturan bisnis dan *machine learning* efektif digunakan sebagai alat bantu pengambilan keputusan di industri teknologi finansial.

Kata Kunci: *Class Weight, Machine Learning, Pinjaman Online, Random Forest, Smote, XGBoost*

ABSTRACT

The growth of online lending services in Indonesia brings significant challenges in the form of late payment and default risks. This study aims to compare the performance of Random Forest and XGBoost algorithms in predicting customer payment status using a historical dataset from Lending Club. The primary challenge in this data is class imbalance, where the number of on-time customers significantly outweighs the default customers. The research methodology includes data preprocessing stages such as data cleaning, handling missing values, normalization using StandardScaler, and addressing class imbalance through SMOTE and Class Weight techniques. The models were optimized using GridSearchCV and Threshold Tuning to obtain an ideal decision threshold. Evaluation results show that the XGBoost + GridSearchCV + Threshold Tuning scenario is the most optimal model, achieving an accuracy of 78.87%, a precision of 58.98%, and an ROC-AUC score of 0.7211. Although Random Forest with Class Weight achieved a higher recall, XGBoost proved superior in maintaining prediction precision and minimizing False Positives. Variables such as interest rate (int_rate), grade, and debt-to-income ratio (dti) were identified as the most significant risk indicators. System implementation through a Streamlit-based prototype demonstrates that integrating business rules with machine learning is effective as a decision support tool in the financial technology industry.

***Keyword**: Class Weight, Machine Learning, Pinjaman Online, Random Forest, Smote, XGBoost*

UBP KARAWANG