

ABSTRAK

Kegagalan kredit nasabah bank merupakan masalah yang cukup serius dalam dunia perbankan. Risiko kredit bermasalah merupakan tantangan utama dalam manajemen kredit. Hal tersebut akan muncul ketika peminjam mengalami kesulitan membayar kembali pinjaman sesuai dengan perjanjian, sehingga ada nasabah yang berhasil membayar kredit namun ada juga yang gagal. Maka dari itu akan dilakukan proses klasifikasi menggunakan algoritma *XGBoost* dalam menangani permasalahan tersebut.

Penelitian ini akan menggunakan data yang di ambil dari *Kaggle* (<https://www.kaggle.com/datasets/nikhil1e9/loan-default/data>). Penelitian ini juga memiliki tujuan utama yaitu untuk mengimplementasikan algoritma *XGBoost* sebagai Pembangunan model awal dan model yang diperbaiki menggunakan *hyperparameter tuning* untuk klasifikasi kegagalan pembayaran kredit nasabah bank. Pengujian dilakukan menggunakan metode *k-fold cross validation* dengan nilai $K = 3, 5, \text{ dan } 10$. Penelitian ini menggunakan proses seleksi fitur berdasarkan *information gain* sebagai metode seleksi fitur yang dapat membantu dalam membuat model yang lebih efisien dan efektif dengan fokus pada fitur-fitur yang paling relevan. Pada proses penggunaan *information gain* nilai batas bawah sudah ditentukan yaitu 0.002, 0.004, 0.006 dan 0.01.

Hasil penelitian menunjukkan bahwa algoritma *XGBoost* memiliki akurasi yang cukup baik mencapai 93.10% dengan menggunakan 7 *hyperparameter tuning*, 14 atribut yang paling relevan berdasarkan hasil penggunaan *information gain* dengan *k-fold* 10 setelah data dikenai proses *balancing* menggunakan teknik SMOTE terhadap data latih dan 88.46% terhadap data uji yang telah melalui tahap *preprocessing* dan juga penggunaan *information gain* sebagai kriteria seleksi fitur dan dengan menggunakan 7 nilai *hyperparameter tuning* terbaik. Hal ini mengindikasikan bahwa *XGBoost* efektif digunakan dalam klasifikasi kegagalan pembayaran kredit nasabah bank ketika data sudah seimbang dan melalui beberapa proses seperti penggunaan *information gain* untuk menentukan atribut yang paling informatif atau relevan dalam memprediksi kelas atau label target, dan juga penggunaan 7 *hyperparameter tuning* yang dapat membantu menaikkan akurasi algoritma.

Kata kunci : *XGBosst*, Kegagalan pembayaran kredit.

ABSTRACT

Credit failure of bank customers is a serious problem in the banking world. Non-performing loan risk is a major challenge in credit management. This will arise when the borrower has difficulty repaying the loan in accordance with the agreement, so that there are customers who succeed in repaying the credit but there are also those who fail. Therefore, the classification process will be carried out using the XGBoost algorithm in handling these problems.

This research will use data taken from Kaggle (<https://www.kaggle.com/datasets/nikhil1e9/loan-default/data>). This research also has the main objective of implementing the XGBoost algorithm as the construction of the initial model and the improved model using hyperparameter tuning for the classification of bank customer credit payment failures. Testing is done using the k-fold cross validation method with a value of $K = 3, 5, \text{ and } 10$. This research uses a feature selection process based on information gain as a feature selection method that can help in creating a more efficient and effective model by focusing on the most relevant features. In the process of using information gain, the lower limit value has been determined, namely 0.002, 0.004, 0.006 and 0.01.

The results showed that the XGBoost algorithm has a fairly good accuracy of 93.10% using 7 hyperparameter tuning, 14 most relevant attributes based on the results of using information gain with k-fold 10 after the data is subjected to a balancing process using the SMOTE technique against training data and 88.46% against test data that has gone through the preprocessing stage and also the use of information gain as a feature selection criterion and by using 7 best hyperparameter tuning values. This indicates that XGBoost is effectively used in the classification of bank customers' credit payment failures when the data is balanced and through several processes such as the use of information gain to determine the most informative or relevant attributes in predicting the target class or label, and also the use of 7 tuning hyperparameters that can help increase the accuracy of the algorithm.

Keywords: XGBosst, Credit payment failure.