

ABSTRAK

Bank syariah memiliki peran penting dalam membangun perekonomian Indonesia. Sayangnya, pangsa bank syariah hanya 6,52% dibandingkan dengan bank lainnya pada September 2021, namun peminat bank syariah mengalami peningkatan yang ditandai dengan meningkatnya jumlah rekening pembiayaan sebanyak 3,1% jika dibandingkan dengan Desember 2020. Dengan adanya perkembangan tersebut, diperlukan konsolidasi validasi nasabah yang menerima pembiayaan untuk memitigasi risiko perusahaan. Untuk itu penelitian ini bertujuan untuk menentukan metode *tree-based* terbaik antara CART, *Random Forest*, dan XGBoost untuk mengklasifikasi kualitas pembiayaan nasabah serta menentukan seberapa signifikan variabel yang digunakan. Berdasarkan teori dan penelitian sebelumnya, XGBoost memberikan hasil yang terbaik dalam mengklasifikasi data yang serupa. Tahapan penelitian dilakukan berdasarkan CRISP-DM, kemudian menggunakan metode SMOTETomek untuk menangani data yang *imbalanced* agar dapat menghasilkan model yang optimal, serta metrik F1-score yang digunakan sebagai metrik perbandingan, kemudian menggunakan RFECV untuk seleksi fitur. Melalui uji coba algoritma dan optimasi parameter, ditemukan bahwa hasil terbaik dihasilkan oleh metode XGBoost dengan F1-score senilai 81,80% dan terdapat 21 variabel yang signifikan terhadap model berdasarkan hasil seleksi variabel pada RFECV. Kata kunci: Klasifikasi, kualitas pembiayaan, CART, *Random Forest*, XGBoost

ABSTRACT

Sharia banks have an important role in building Indonesia's economy. Unfortunately, sharia banks' market share is only 6,52% compared to other banks as per September 2021, however sharia banks have gained popularity which was signified by the 3,1% increase in the number of financing accounts compared to December 2020. With aforementioned growth, consolidation of validating customers for financing is needed to mitigate the company risks. Therefore, the purpose of this research is to determine the best tree-based method between CART, Random Forest, and XGBoost to classify the financing quality of customers and to determine how significant the used variables are. Based on theory and previous research, XGBoost gave the best result in classifying similar datasets. This research's procedure is based on CRISP-DM, then the SMOTETomek method is used to handle imbalanced data in order to produce the best model, and also the F1-score metric is used as the comparison metric, and then using RFECV for feature selection. Through trials and parameter optimization, it is found that the best result came from XGBoost method with F1-score valued at 81.80% and there

are 21 significant variables toward the model based on the results of feature selection on RFECV.

Keywords: Classification, financing quality, CART, Random Forest, XGBoost

