

ABSTRAK

PERBANDINGAN KINERJA MODEL XGBOOST, LIGHTGBM DAN CATBOOST DALAM PREDIKSI *CUSTOMERS CHURN* INDUSTRI PERBANKAN

Oleh

RIVALDO

Customer churn adalah kondisi ketika pelanggan berhenti menggunakan layanan suatu perusahaan, dan dalam industri perbankan hal ini dapat berdampak signifikan terhadap pendapatan dan citra lembaga. Mengidentifikasi pelanggan yang berpotensi *churn* menjadi penting agar bank dapat menyusun strategi retensi yang lebih tepat sasaran. Penelitian ini membandingkan kinerja tiga algoritma *machine learning* berbasis *gradient boosting*, yaitu XGBoost, LightGBM, dan CatBoost, dalam memprediksi *customer churn* menggunakan *dataset* pelanggan kartu kredit. Proses mencakup penyeimbangan data dengan ADASYN, *tuning hyperparameter*, dan validasi menggunakan *Stratified K-Fold Cross Validation*. Evaluasi dilakukan berdasarkan akurasi, *precision*, *recall*, dan *F1-score* pada dua skema pembagian data.

Hasil menunjukkan bahwa XGBoost menghasilkan akurasi tertinggi, yakni 98,3% pada skema pertama dan 96,4% pada skema kedua, diikuti oleh LightGBM (98,1% dan 95,3%) serta CatBoost (97,1% dan 96,4%). Hal ini menunjukkan bahwa XGBoost lebih unggul dalam menangani data *churn* perbankan dibanding dua model lainnya. Model XGBoost direkomendasikan sebagai model yang paling optimal untuk diterapkan pada kasus serupa di sektor perbankan.

Kata Kunci: *Customer Churn*, XGBoost, LightGBM, CatBoost, *Machine Learning*.

ABSTRACT

COMPARISON OF THE PERFORMANCE OF XGBOOST, LIGHTGBM, AND CATBOOST MODELS IN PREDICTING CUSTOMER CHURN IN THE BANKING INDUSTRY

By

RIVALDO

Customer churn refers to the condition where customers stop using a company's services, which can significantly impact revenue and institutional reputation in the banking industry. Identifying customers who are likely to churn is crucial for banks to develop more targeted retention strategies. This study compares the performance of three gradient boosting-based machine learning algorithms XGBoost, LightGBM, and CatBoost in predicting customer churn using a credit card customer dataset. The process includes data balancing with ADASYN, hyperparameter tuning, and validation using Stratified K-Fold Cross Validation. Evaluation is based on accuracy, precision, recall, and F1-score using two data splitting schemes.

The results show that XGBoost achieved the highest accuracy, namely 98.3% in the first scheme and 96.4% in the second, followed by LightGBM (98.1% and 95.3%) and CatBoost (97.1% and 96.4%). These results indicate that while XGBoost performs well in handling banking churn data, XGBoost is the most optimal model to be applied in similar use cases in the banking sector.

Keywords: *Customer Churn, XGBoost, LightGBM, CatBoost, Machine Learning.*