

## ABSTRAK

### PERBANDINGAN TEKNIK OVERSAMPLING DALAM MENGATASI PERMASALAHAN KETIDAKSEIMBANGAN KELAS PADA KASUS PENIPUAN KARTU KREDIT

Oleh

DEAN FARHAN LAZUARDI

9882405120111009

Peningkatan jumlah pengguna kartu kredit di Indonesia sejalan dengan peningkatan tingkat kejahatan di sektor perbankan, khususnya penipuan transaksi kartu kredit. Algoritma klasifikasi pada *machine learning* dapat diimplementasikan untuk mengklasifikasikan penipuan transaksi kartu kredit. Namun, implementasi model *machine learning* dalam klasifikasi penipuan transaksi kartu kredit juga menghadapi banyak tantangan, terutama terkait dengan ketidakseimbangan kelas. Untuk mengatasi masalah tersebut maka digunakan Teknik oversampling. Pada penelitian ini akan diukur kinerja teknik oversampling dalam menyeimbangkan jumlah kelas serta mengukur dan membandingkan efek setiap teknik oversampling terhadap performa model klasifikasi. Teknik *oversampling* yang digunakan adalah *SMOTE*, *Borderline SMOTE*, dan *ADASYN*. Adapun algoritma klasifikasi yang digunakan adalah *Decision Tree* dan *Logistic Regression*. Data yang digunakan dalam penelitian ini adalah *card fraud detection* yang didapatkan dari Kaggle.com. Data yang digunakan berjumlah 284.807 baris data, memiliki 284.315 data kelas *genuine* dan 492 data kelas *fraud*, menunjukkan ketidakseimbangan kelas yang sangat signifikan dengan rasio 575:1. Langkah yang dilakukan dalam penelitian ini adalah melakukan *EDA*, *preprocessing* data, *hyperparameter tuning*, pelatihan model, dan pengujian model. Hasilnya Teknik *oversampling* yang digunakan berhasil menyeimbangkan jumlah kelas pada dataset, tetapi cenderung mengurangi performa model terutama pada presisi dan F1 score terutama pada *Logistic Regression* yang mengalami penurunan signifikan. Tetapi Teknik *oversampling* berhasil menaikkan *AUROC* pada mayoritas model dan kenaikan recall yang signifikan pada model *Logistic Regression*.

Kata kunci: penipuan, transaksi kartu kredit, ketidakseimbangan kelas, klasifikasi, oversampling

## **ABSTRACT**

### **COMPARISON OF OVERSAMPLING TECHNIQUES IN ADDRESSING CLASS IMBALANCE IN CREDIT CARD FRAUD CASES**

*By*

**DEAN FARHAN LAZUARDI**

**9882405120111009**

*The increasing number of credit card users in Indonesia is in line with the increasing level of crime in the banking sector, especially credit card transaction fraud. Machine learning classification algorithms can be implemented to classify credit card transaction fraud. However, the implementation of machine learning models in credit card transaction fraud classification also faces many challenges, especially related to class imbalance. To overcome this problem, oversampling technique is used. This research will measure the performance of oversampling techniques in balancing the number of classes and measure and compare the effect of each oversampling technique on the performance of the classification model. The oversampling techniques used are SMOTE, Borderline SMOTE, and ADASYN. The classification algorithms used are Decision Tree and Logistic Regression. The data used in this research is card fraud detection obtained from Kaggle.com. The data used amounted to 284,807 rows of data, having 284,315 genuine class data and 492 fraud class data, showing a very significant class imbalance with a ratio of 575:1. The steps taken in this research are conducting EDA, data preprocessing, hyperparameter tuning, model training, and model testing. The oversampling technique used succeeded in balancing the number of classes in the dataset, but tended to reduce model performance, especially in precision and F1 score, especially in Logistic Regression which experienced a significant decline. However, the oversampling technique managed to increase AUROCC in the majority of models and a significant increase in recall in the Logistic Regression model.*

*Keywords: fraud, credit card transactions, class imbalance, classification, oversampling*