ABSTRAK

Dharmana, I Wayan (2024), *Klasifikasi Transaksi Fraud Kartu Kredit Menggunakan Oversampling ADASYN dan Seleksi Fitur SVM-RFECV*. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I: Dr. I Gede Aris Gunadi, S.Si. M.Kom. dan Pembimbing II: Dr. Luh Joni Erawati Dewi, S.T., M.Pd.

Kata Kunci: kartu kredit, klasifikasi *fraud*, *machine learning, imbalanced dataset*

Perkembangan kejahatan transaksi *fraud* kartu kredit memberikan dampak kerugian finansial bagi pemegang kartu. Pengembangan model prediksi transaksi *fraud* menggunakan *machine learning* telah dilakukan namun memiliki beberapa tantangan meliputi ketidakseimbangan data serta dimensi dataset yang cukup besar. Penelitian ini mengusulkan pendekatan dengan seleksi fitur dengan SVM-RFECV dan metode *oversampling* menggunakan ADASYN. Pembagian data latih dilakukan dengan beberapa rasio data latih meliputi 70%, 80% dan 90%. Data latih yang telah dibagi dilakukan seleksi fitur untuk mendapatkan sejumlah variabel paling optimal sesuai kontribusi terhadap variabel dependen (*isFraud)*. Seleksi fitur menghasilkan variabel optimal dengan rincian 390 variabel pada data latih 70%, 400 variabel pada data latih 80%, dan 390 variabel pada data latih 90%. Tahapan *oversampling* dengan ADASYN dilakukan pada strategi *oversampling* 100%, 50% dan 25%. Data latih yang telah dilakukan *oversampling* dilakukan analisis uji kualitas data menggunakan histogram distribusi dan *silhouette index*. Uji klasifikasi dilakukan menggunakan algoritma AdaBoost dan LGBM dengan metrik kinerja AUC, *precision*, *recall* dan *f1 score*. Berdasarkan analisis dan evaluasi strategi *oversampling*, strategi dengan data latih 80%, data uji 20%, rasio *oversampling* 100% dengan algoritma LGBM menjadi strategi yang direkomendasikan berdasarkan analisis kualitas data dan kinerja AUC yang optimal sejumlah 86,27%.

ABSTRACT

Dharmana, I Wayan (2024), *Classification of Credit Card Fraud Transactions Using ADASYN Oversampling and SVM-RFECV Feature Selection.* Thesis, Computer Science, Postgraduate Program, Ganesha University of Education.

This thesis has been supervised and approved by Supervisor I: Dr. I Gede Aris Gunadi, S.Si. M.Kom. and Supervisor II: Dr. Luh Joni Erawati Dewi, S.T., M.Pd.

Keywords: *credit card, fraud classification, machine learning, imbalanced dataset*

The growth of credit card fraud transaction crimes has resulted in financial losses for card holders. The development of fraud transaction prediction model using machine learning has been carried out but has several challenges including data imbalance and large dataset dimensions. This research proposes an approach with feature selection using SVM-RFECV and oversampling method using ADASYN. The distribution of training data is carried out with several training data ratios including 70%, 80% and 90%. The training data that has been divided is subjected to feature selection to obtain the most optimal number of variables according to the contribution to the dependent variable (isFraud). Feature selection produces optimal variables with details of 390 variables on 70% training data, 400 variables on 80% training data, and 390 variables on 90% training data. The oversampling stages with ADASYN are carried out at 100%, 50% and 25% oversampling strategies. Training data that has been oversampled is subjected to data quality analysis using distribution histograms and silhouette index. The classification test was carried out using the AdaBoost and LGBM algorithms with the performance metrics AUC, precision, recall and f1 score. Based on the analysis and evaluation of oversampling strategies, a strategy with 80% training data, 20% testing data, 100% oversampling ratio, and LGBM algorithm is the recommended strategy based on data quality analysis and optimal AUC performance of 86,27%.