

## **ABSTRACT**

*Identification of financial transactions as suspicious or fraudulent transactions that are indicated as money laundering is mostly done manually so that it is not optimal. Data mining techniques can be a solution to overcome the limitations of the manual method. The main challenge in applying data mining techniques for financial fraud detection is an imbalanced dataset with the proportion of fraud class is much smaller than non-fraud. This causes the model to produce unbalanced precision and recall, resulting in a low f1-score. It means that the model can predict one class well, but not with another class. In addition, the need for high computational resources in implementing data mining for high-dimensional data and the duration of model training also makes it a special consideration.*

*This study intends to improve the performance of fraud detection systems in the financial sector that is indicated by suspicious financial transactions. In this study, the approach to fraud detection in financial transactions is carried out with data mining technique using tuned classifier based on Support Vector Machine (SVM). The classifier performance improvement is performed using SVM tuning by adjusting the kernel and hyperparameter sets combined with the Random Under Sampling (RUS) technique and the modified approach of the Principal Component Analysis (mPCA). Specifically, RUS is used to handle imbalanced datasets and cuts model training time. While mPCA modifies data dimension reduction techniques based on classification errors to speed up computational time without disturbing the performance of SVM.*

*As the result, with this combination technique, the classifier can detect fraud more accurately with an increase in precision of 29.31%, recall 2% and f1-score of 19.8% compared to the previous study.*

**Keywords:** Money Laundering, Fraud Detection, SVM, RUS, PCA, Data Mining

## INTISARI

Identifikasi transaksi keuangan mencurigakan yang terindikasi sebagai tindak pidana pencucian uang saat ini masih dilakukan secara manual sehingga belum optimal. Teknik *data mining* dapat menjadi solusi untuk mengatasi masalah keterbatasan cara manual. Tantangan utama dalam menerapkan teknik *data mining* untuk deteksi *fraudulent transaction* adalah *imbalanced dataset*, yakni proporsi data kelas *fraud* yang jauh lebih kecil daripada *non-fraud*. Hal ini menyebabkan model menghasilkan *precision* dan *recall* yang tidak seimbang, sehingga *f1-score* menjadi rendah. Akibatnya, model dapat memprediksi satu kelas dengan baik, tetapi tidak dengan kelas yang lain. Selain itu, kebutuhan sumber daya komputasi yang besar dalam penerapan *data mining* pada data berdimensi tinggi serta lamanya waktu pelatihan model juga menjadi perhatian tersendiri.

Penelitian ini bermaksud melakukan peningkatan kinerja sistem deteksi *fraud* dalam bidang keuangan yang diindikasikan dengan adanya transaksi keuangan mencurigakan. Pada penelitian ini, pendekatan deteksi *fraud* dilakukan dengan teknik *data mining* menggunakan *tuned classifier* berbasis *Support Vector Machine (SVM)*. Peningkatan kinerja *classifier* dilakukan dengan *tuning* pada SVM dengan mengatur *kernel* dan *hyperparameter* yang dikombinasikan dengan teknik *Random Under Sampling (RUS)* dan *modified approach of Principal Component Analysis (mPCA)*. Secara khusus, RUS digunakan untuk menangani *imbalanced dataset* dan memangkas waktu pelatihan model. Sedangkan mPCA memodifikasi teknik reduksi dimensi data dengan berbasiskan *classification error* untuk mempercepat waktu komputasi tanpa mengganggu kinerja SVM.

Hasilnya, dengan teknik kombinasi ini, *classifier* dapat melakukan deteksi *fraud* lebih akurat dengan peningkatan *precision* sebesar 29.31%, *recall* 2% dan *f1-score* 19.8% dibandingkan dengan penelitian sebelumnya.

**Kata kunci** -- Pencucian Uang, Deteksi *Fraud*, SVM, RUS, PCA, *Data Mining*.