

1. PENDAHULUAN

1.1. Latar Belakang

Pada era digitalisasi, pengembangan Transaksi yang mencurigakan di rekening bank tidak hanya membawa risiko besar bagi industri perbankan, tetapi juga berdampak serius terhadap kepentingan pedagang dan nasabah bank, yang menyebabkan kerugian langsung maupun tidak langsung yang signifikan dengan tingkat yang bervariasi[1]. Upaya pencegahan terhadap kecurangan transaksi bank terus berkembang. Salah satu langkah pencegahan terhadap kecurangan tersebut dengan upaya pendeteksian pola-pola yang mencurigakan (*fraud*) dari suatu transaksi yang disebut juga *fraud detection*. *Fraud detection* sering kali melibatkan pendekatan *machine learning* untuk membangun model yang dapat mendeteksi antara transaksi normal(*non-fraud*) dan transaksi mencurigakan (*fraud*).

Penelitian terkait *Fraud Detection* sudah banyak dilakukan, salah satunya oleh Muhammad Sopiyan, dkk. yang menunjukkan *Random Forest Classifier* (RFC) memiliki performansi terbaik dibanding dengan algoritma lain dengan nilai akurasi pada data *training* adalah 100% dan data *testing* 99.99% [2]. Hasil serupa juga didapat oleh K. Deepika, dkk. dengan hasil pada nilai *accuracy* 100%, *recall* 78%, *precision* 91%, dan *F1-Score* 84% [3].

Berdasarkan beberapa penelitian yang telah dilakukan, akan dilakukan deteksi kecurangan menggunakan algoritma *Random Forest* terhadap dataset transaksi bank yang telah diolah oleh aplikasi *Solve Ease Fraud Investigation System* (SEFIS) dalam jangka waktu 1 November 2023 hingga 19 November 2023. Penelitian ini menggunakan algoritma *Random Forest* karena dari beberapa kasus yang telah dilakukan oleh peneliti sebelumnya dengan melakukan komparasi terhadap beberapa algoritma, hasil menunjukkan algoritma *Random Forest* dapat mengelola permasalahan *fraud detection* dengan baik.

Pada dataset transaksi yang akan kami gunakan memiliki karakteristik data yang tidak seimbang (*imbalance*). Ketidakeimbangan(*imbalance*) data terlihat dengan data transaksi normal(*non-fraud*) memiliki nilai mayoritas dan data transaksi

mencurigakan (*fraud*) memiliki nilai minoritas. Perlu dilakukan penyeimbangan data (*Balancing Data*) dengan menerapkan *undersampling* terhadap data mayoritas. Untuk mengetahui performa yang baik, kami melakukan ujicoba dengan melakukan komparasi terhadap data yang dilakukan *undersampling* dan data yang tidak dilakukan *undersampling* dengan menganalisa dengan beberapa metrik performansi yang meliputi: a. *Accuracy*, b. *Precision*, c. *Recall*, d. *Specificity*, dan e. *F1-Score*.

kontribusi utama pada penelitian ini adalah penerapan *machine learning* dengan model algoritma *Random Forest* pada data transaksi bank khususnya di Indonesia untuk mendeteksi kecurangan yang terjadi dengan secara menyeluruh. Dengan melakukan perbandingan data asli tanpa *Balancing* dan data yang dilakukan *Balancing* menggunakan *Random Undersampling (RUS)*, penelitian ini memberikan wawasan terhadap pendeteksian pada kecurangan transaksi bank menggunakan *machine learning*, khususnya pada algoritma *Random Forest*, serta dapat dilakukan implementasi pada aplikasi *Solve Ease Fraud Investigation System (SEFIS)* untuk pengembangan sistem deteksi kecurangan menggunakan *machine learning* dengan model algoritma *Random Forest* di masa depan.

1.2. Perumusan Masalah

Berdasarkan latar belakang di atas, maka dirumuskan permasalahan yang diangkat sebagai berikut:

1. Bagaimana membangun *fraud detection system* menggunakan algoritma *random forest*?
2. Bagaimana kinerja algoritma *random forest* pada *fraud detection system* dalam memprediksi transaksi yang termasuk *fraud* atau *non-fraud*?

1.3. Batasan Masalah

Batasan permasalahan pada penelitian tugas akhir ini mencakup:

1. *Fraud* pada kasus ini merupakan data yang dikategorikan sebagai *suspected fraud* yang dilakukan oleh aplikasi *Solve Ease Fraud Investigation System (SEFIS)*

1.4. Tujuan

Tujuan yang di capai pada penelitian tugas akhir ini, yaitu:

1. Membangun *Fraud Detection System* menggunakan metode *Random Forest*.
2. Mengukur kinerja algoritma *Random Forest* untuk *Fraud Detection System* dengan metrik performansi *accuracy*, *precision*, *recall*, *specificity* dan *F1-Score*

1.5. Sistematika Penulisan

Dalam laporan ini, sistematika penulisan dibagi menjadi lima bab yang dijabarkan sebagai berikut:

1. BAB I Pendahuluan

Pada BAB I dijelaskan terkait permasalahan *fraud detection*, metode *Random Forest*, dan dataset secara umum terkait kasus yang akan dibahas.

2. BAB II Kajian Pustaka

Pada BAB II dijelaskan terkait dasar-dasar teori mengenai *Fraud Detection*, *Random Forest*, dan *Confusion Matrix*.

3. BAB III Perancangan Sistem

Pada BAB III dijelaskan mengenai alur sistem yang dirancang terdiri dari *Preprocessing data*, *Splitting Dataset*, dan *Hyperparameter Tuning*.

4. BAB IV Pengujian dan Analisis

Pada BAB IV dilakukan pemaparan hasil dan analisa dari pengujian yang telah dilakukan.

5. BAB V Kesimpulan dan Saran

Pada BAB V diambil Kesimpulan dari hasil Analisa serta pengujian yang telah dilakukan serta saran untuk penelitian selanjutnya.