

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan yang cepat dari bisnis online di Indonesia mencerminkan pengakuan akan era di mana pemanfaatan teknologi informasi semakin diakui keberadaannya. E-commerce, atau bisnis online, semakin berkembang pesat di Indonesia karena pertumbuhan internet yang signifikan serta perubahan perilaku konsumen. Akses yang mudah ke internet melalui wifi atau perangkat gadget telah memberikan kenyamanan bagi masyarakat untuk mencari informasi tentang produk atau layanan yang mereka perlukan. Ditambah dengan upaya promosi yang agresif dari perusahaan e-commerce yang menawarkan berbagai fasilitas, semakin memudahkan masyarakat dalam memperoleh barang atau layanan yang mereka inginkan (Sidharta et al., 2015).

Transaksi dalam bisnis online, terutama dalam aktivitas jual-beli, menunjukkan sejumlah kemudahan, termasuk upaya promosi yang lebih mudah, kelancaran dalam transaksi, pengurangan biaya, percepatan proses transaksi, peningkatan penjualan, peningkatan kinerja perusahaan, dan akhirnya, penekanan pada efisiensi serta efektivitas sebagai alasan utama (Djuhari & Dewi, 2021). Penggunaan transaksi online telah menjadi suatu keharusan dalam rutinitas sehari-hari. Namun, kemudahan dan kenyamanan dalam melakukan transaksi online juga membawa risiko penipuan dan kejahatan ke dunia digital. Untuk melindungi konsumen dan bisnis online dari penipuan, diperlukan analisis data fraud untuk mengidentifikasi transaksi yang mencurigakan atau potensial penipuan.

Meningkatnya penggunaan transaksi online, kasus fraud pada transaksi online pun semakin banyak. Fraud adalah tindakan yang disengaja untuk memperoleh keuntungan atau tujuan tertentu. Suatu tindakan dikategorikan sebagai fraud jika melibatkan korban, dilakukan dengan sengaja, korban menuruti keinginan pelaku, dan mengakibatkan kerugian pada korban. Berdasarkan informasi dari bank Mandiri, peningkatan kasus kejahatan finansial terkait dengan melemahnya kondisi ekonomi. (Armiani & Agustini, 2022).

Penipuan telah ada sejak awal mula *e-commerce*. Namun, lonjakan penjualan dan pembelian online akibat pandemi COVID-19 memberikan peluang baru bagi penipu. Pada tahun 2020, penipuan belanja online menyumbang 38 persen dari seluruh penipuan yang dilaporkan di seluruh dunia, naik dari 24 persen sebelum wabah ini terjadi. Meskipun angka tersebut telah menurun sejak krisis mereda, pelanggaran keamanan terus memberikan dampak besar pada industri ini, yang kerugian akibat penipuan pembayaran online berjumlah lebih dari

40 miliar dolar AS pada tahun 2022. Mengingat situasi ini, penipuan e-commerce pasar deteksi dan pencegahan diperkirakan akan tumbuh lebih dari dua kali lipat antara tahun 2023 dan 2027, melebihi 100 miliar dolar.

Untuk menghadapi permasalahan *fraud* di industri perbankan, diperlukan langkah-langkah dan solusi yang efektif. Salah satu pendekatan yang bisa digunakan adalah analisis *fraud* menggunakan teknik *machine learning*. *Machine learning* merupakan bagian dari bidang kecerdasan buatan (*artificial intelligence*) yang merujuk pada program komputer dengan *algoritma* yang memungkinkannya untuk belajar dari data dan bertindak mirip dengan cara manusia. Dalam kerangka kecerdasan buatan, terdapat beberapa metode, dan salah satunya adalah *machine learning*. Metode ini melibatkan penggunaan program-program yang diprogram sedemikian rupa sehingga komputer dapat mengadopsi tingkah laku cerdas mirip manusia, meningkatkan pemahaman dan pengalaman secara otomatis (Armiani & Agustini, 2022).

Mengatasi masalah tersebut dapat diselesaikan dengan menerapkan teknik *data mining* yang disebut klasifikasi. Tujuan dari klasifikasi yaitu untuk memprediksi label atau kelas dari suatu objek berdasarkan informasi atau atribut yang tersedia. Metode yang termasuk kedalam klasifikasi diantaranya yaitu metode *random forest*. *Random forest* mampu mengatasi masalah non-linier karena didasarkan pada teknik pohon keputusan (*decision tree*). Konsep dasar *random forest* yaitu menggunakan lebih dari classifier dari metode yang sama kemudian mengkombinasikannya melalui voting untuk mendapatkan hasil dugaan klasifikasi akhir. *Random forest* memiliki kelebihan yaitu diantaranya dapat memberikan hasil klasifikasi yang baik disertai dengan hasil residu/error yang lebih rendah, dapat mengatasi data training yang berukuran sangat besar secara efisien, selain itu metode *random forest* juga dinilai efektif dalam mengatasi masalah missing data (Lestari & Sirodj, 2022).

Algoritma Random Forest adalah sebuah metode dalam machine learning yang sangat efisien dalam mengklasifikasikan dan melakukan prediksi terhadap dataset yang besar, termasuk pada dataset yang digunakan dalam analisis data fraud. Dengan kemampuan untuk mengidentifikasi pola dan relasi dalam data secara otomatis, algoritma *Random Forest* dapat digunakan untuk mengklasifikasikan transaksi sebagai fraud atau tidak fraud. Dalam analisa data fraud pada transaksi online, Algoritma *Random Forest* dapat digunakan untuk membangun model prediksi yang mampu mengidentifikasi pola atau perilaku transaksi yang mencurigakan, sehingga transaksi fraud dapat diidentifikasi dan ditindaklanjuti secara tepat waktu. Dengan demikian, analisis data fraud pada transaksi online menggunakan algoritma *Random Forest* sangat penting untuk meningkatkan keamanan dan kenyamanan dalam bertransaksi online bagi konsumen dan bisnis online.

Pendekatan data driven sangat penting diterapkan dalam klasifikasi menggunakan algoritma *random forest*, karena sebagai algoritma *machine learning random forest* di desain untuk mempelajari pola dan tren pada data untuk memprediksi kelas atau label yang tepat dari suatu data baru. Pendekatan data driven untuk mengoptimalkan performa model klasifikasi menggunakan algoritma *random forest*, dilakukan beberapa tahapan seperti pemilihan fitur yang relevan, pemilihan parameter yang optimal, serta evaluasi pemeliharaan model yang terus menerus. Dalam pemilihan fitur, data driven dapat digunakan untuk menentukan fitur yang paling berpengaruh dalam klasifikasi dengan menganalisis korelasi antara fitur-fitur dan label kelas. Hal ini akan membantu mengurangi dimensi data dan meningkatkan akurasi model.

Selain itu, data driven juga diperlukan dalam pemilihan parameter algoritma, seperti jumlah pohon keputusan dan kriteria pemilihan fitur yang digunakan. Dengan melakukan eksperimen menggunakan data pelatihan dan pengujian, kita dapat menentukan parameter terbaik yang dapat menghasilkan model klasifikasi dengan performa yang optimal.

Data driven juga sangat penting dalam evaluasi dan pemeliharaan model klasifikasi *random forest*. Evaluasi dilakukan dengan menguji performa model pada data pengujian, dan memastikan bahwa model memiliki performa yang baik pada data pengujian, dan memastikan bahwa model memiliki performa yang baik pada data yang belum pernah dilihat sebelumnya. Pemeliharaan dilakukan dengan memperbarui model menggunakan data yang baru, sehingga model dapat terus berkembang dan mempertahankan performanya.

Dengan demikian, data driven sangat penting dalam penerapan klasifikasi menggunakan algoritma *random forest* untuk meningkatkan akurasi model dan menghasilkan prediksi yang lebih akurat.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka dapat diidentifikasi permasalahan-permasalahan diantaranya:

1. Dalam era digital, transaksi online semakin banyak dilakukan oleh masyarakat. Namun, hal ini juga meningkatkan resiko fraud pada transaksi online. Oleh karena itu, penting untuk melakukan analisis fraud pada transaksi online untuk mengurangi resiko kerugian.
2. Metode analisis fraud tradisional seperti rule-based dan model statistik memiliki keterbatasan dalam menganalisis fraud pada transaksi online yang semakin kompleks. Oleh karena itu, diperlukan metode deteksi fraud yang lebih canggih dan dapat mengatasi kompleksitas transaksi online.

3. Data transaksi online sangat kompleks karena terdiri dari berbagai variabel seperti set, type, amount, nameOrig, oldbalanceOrg, nameDest, dan oldbalanceDest. Oleh karena itu, diperlukan pendekatan data driven untuk mengelola dan menganalisis data transaksi online dengan baik.
4. Algoritma Random Forest bisa digunakan untuk mengidentifikasi penipuan dalam transaksi online. Namun, untuk mendapatkan hasil yang tepat dan akurat, performa dari algoritma ini perlu ditingkatkan melalui optimisasi yang sesuai.

1.3 Ruang Lingkup Masalah

Ruang lingkup pembahasan dalam penelitian ini dibatasi pada penerapan algoritma *random forest* yang digunakan untuk menganalisis fraud pada transaksi online. Kemudian dilakukan Indikator penilaian untuk mengevaluasi kinerja metode pada *machine learning* menggunakan *confusion matrix*. Sumber data yang digunakan yaitu dataset kaagle. Sedangkan *software* yang digunakan adalah *Jupyter Notebook* dan Bahasa pemrograman yang digunakan yaitu bahasa *python*.

1.4 Rumusan Masalah

Berdasarkan latar belakang diatas dapat di simpulkan rumusan masalah sebagai berikut:

1. Bagaimana menerapkan algoritma *random forest* untuk menganalisis faktor faktor yang menyebabkan *fraud* pada transaksi online?
2. Bagaimana hasil tingkat akurasi dari analisis transaksi online yang bersifat *fraud* menggunakan metode *confussion matrix*?

1.5 Batasan Masalah

Untuk memfokuskan penelitian dan mencegah perluasan topik, peneliti merumuskan batasan permasalahan sebagai berikut:

1. Data yang digunakan dalam penelitian adalah data transaksi online yang di dapat dari situs kaagle
2. Variabel yang digunakan dalam penelitian ini yaitu set, type, amount, nameOrig, oldbalanceOrg, newbalanceOrig, nameDest, dan oldbalanceDest.
3. Metode yang digunakan dalam menganalisis fraud pada transaksi online adalah metode *random forest*.

1.6 Tujuan Penelitian

Berdasarkan perumusan masalah yang ada, maka yang menjadi tujuan penelitian ini adalah:

1. Untuk mengetahui hasil penerapan algoritma *random forest* untuk menganalisis faktor faktor yang menyebabkan fraud pada transaksi online
2. Untuk mengetahui hasil tingkat akurasi dari analisis transaksi online yang bersifat fraud menggunakan metode *confussion matrix*

1.7 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Bagi penulis, memberikan tambahan pengetahuan atas penggunaan algoritma *random forest* dalam mendeteksi fraud pada transaksi online
2. Bagi pembaca, berguna sebagai referensi atau rujukan dalam membuat penelitian selanjutnya yang terkaita dengan masalah yang serupa.