

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dalam era digitalisasi, aktivitas pencatatan dan pengelolaan transaksi keuangan semakin kompleks dan terotomatisasi. Baik di sektor bisnis, pemerintahan, maupun organisasi sosial, sistem keuangan memproses ribuan hingga jutaan transaksi setiap hari. Namun, kemajuan ini juga membuka celah baru terhadap potensi anomali transaksi seperti kesalahan pencatatan, penggandaan transaksi, hingga penipuan finansial (*anomali*). Deteksi anomali dalam transaksi sangat penting dalam menjaga integritas sistem keuangan *modern*, terutama pada sistem pembayaran elektronik yang rentan terhadap penyalahgunaan. (Waspada dkk., 2020)

Dengan jumlah data yang besar dan beragam, sangat sulit untuk mendeteksi anomali transaksi secara manual. Sistem keuangan yang tidak memiliki deteksi otomatis berisiko gagal mengidentifikasi transaksi mencurigakan secara cepat. Penelitian dari (Gillespie, 2019) menunjukkan bahwa organisasi seringkali menghadapi tantangan dalam mengenali *outlier* karena ketidakteraturan distribusi data dan kurangnya data label sebagai pembanding.

Selain itu, data transaksi keuangan pada umumnya memiliki karakteristik tidak seimbang (*imbalanced*), dimana jumlah transaksi normal jauh lebih banyak dibandingkan transaksi anomali. Kondisi ini menyebabkan proses deteksi menjadi lebih menantang karena model dapat cenderung mengklasifikasikan sebagian besar data sebagai transaksi normal. Oleh karena itu, diperlukan metode yang mampu mendeteksi anomali secara efektif meskipun proporsi data anomali sangat kecil (Ounacer dkk., 2018).

Untuk menjawab permasalahan tersebut, dibutuhkan suatu sistem cerdas yang mampu mendeteksi anomali secara otomatis dan akurat. Sistem ini juga perlu memiliki antarmuka (*dashboard*) yang dapat memvisualisasikan hasil deteksi secara interaktif, sehingga memudahkan pengguna dalam memahami dan mengambil keputusan berdasarkan data. Untuk mendapatkan hasil yang maksimal, diperlukan metode yang dapat mendeteksi anomali tersebut secara akurat.

Selain pemilihan metode, kualitas data juga berpengaruh terhadap performa model machine learning. Keberadaan data duplikat dapat menyebabkan distribusi data menjadi kurang representatif dan berpotensi memengaruhi hasil prediksi yang dihasilkan oleh model. Oleh karena itu, perlu dilakukan analisis mengenai pengaruh keberadaan data duplikat terhadap kinerja algoritma Isolation Forest dalam mendeteksi transaksi anomali.

Kinerja algoritma deteksi anomali tidak hanya dipengaruhi oleh metode yang digunakan, tetapi juga oleh konfigurasi parameter yang diterapkan. Parameter seperti jumlah pohon isolasi dan *contamination rate* dapat memengaruhi kemampuan model dalam membedakan transaksi normal dan anomali. Oleh karena itu, diperlukan proses penentuan parameter yang optimal agar model dapat menghasilkan performa deteksi yang lebih baik (John dkk., 2019).

Metode *Isolation Forest* merupakan salah satu algoritma yang bisa diimplementasikan dalam sistem ini. Algoritma *Isolation Forest* ini berbasis pembentukan pohon acak yang mengisolasi data *outlier*. Cara kerjanya yaitu mempelajari pola mayoritas kemudian mengisolasi data secara acak, namun berdasarkan transaksi yang berbeda dari mayoritas. Berdasarkan penelitian oleh (Kumar dkk., 2021) metode ini terbukti efektif dalam mendeteksi *anomali* pada transaksi kartu pembayaran tanpa perlu pelabelan data sebelumnya.

Melalui pengembangan *dasbord* dengan menerapkan metode *Isolation Forest* ini, diharapkan dapat membantu instansi atau organisasi mendeteksi transaksi keuangan yang mencurigakan. Pendekatan ini tidak hanya membantu dalam menemukan anomali secara otomatis, namun juga memberikan gambaran visual yang memudahkan pengguna dalam pengambilan keputusan berbasis data. Sistem ini juga dapat membuat pekerjaan jadi lebih cepat, akurat dan efisien.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana parameter tuning yang optimal seperti jumlah *estimators*, *contamination rate* terhadap performa model *Isolation Forest* dalam mendeteksi anomali transaksi?

2. Bagaimana kualitas data dapat mempengaruhi kinerja model *Isolation Forest* dalam mendeteksi anomali?
3. Bagaimana kinerja metode *Isolation Forest* dalam mengidentifikasi anomali transaksi?

### 1.3 Tujuan

1. Mencari Parameter tuning yang optimal model *Isolation Forest*, seperti jumlah *estimators* dan *contamination* terhadap performa deteksi anomali.
2. Mengetahui pengaruh kualitas data terhadap kinerja model *Isolation Forest* dalam mendeteksi anomali pada data transaksi keuangan.
3. Mengevaluasi performa dan kinerja metode *Isolation Forest* terhadap data keuangan yang memiliki karakteristik tidak seimbang.

### 1.4 Manfaat

Memberikan solusi praktis berupa sistem *dashboard* yang dapat digunakan untuk memantau transaksi secara otomatis serta membantu suatu instansi atau organisasi dalam mendeteksi adanya transaksi mencurigakan secara efisien. Selain itu, sistem ini juga memberikan pemahaman tentang pentingnya kualitas data dalam proses deteksi anomali menggunakan *Isolation Forest*. Serta Memberikan informasi mengenai pengaruh tuning parameter pada model *Isolation Forest* terhadap performa deteksi

### 1.5 Batasan Masalah

Untuk menjaga fokus dan ruang lingkup penelitian, batasan dalam penelitian ini adalah:

1. Data transaksi yang digunakan yaitu *Dataset* bersumber dari kaggle (<https://www.kaggle.com/code/gpreda/credit-card-fraud-detection-predictive-models/>)
2. Visualisasi hasil hanya difokuskan pada *dashboard* interaktif berbasis *Python (Streamlit)* dan tidak melibatkan sistem keamanan atau integrasi API eksternal.
3. Penelitian ini hanya membahas proses deteksi anomali transaksi.
4. Kualitas data diukur berdasarkan data duplikat.
5. Tidak membandingkan dengan metode yang lain.