BAB I
PENDAHULUAN

1.1 Latar Belakang

Era keuangan digital telah membawa kemajuan signifikan dalam transaksi elektronik, namun juga memunculkan ancaman serius berupa penipuan keuangan (*financial fraud*) yang semakin canggih dan merugikan. Kecurangan keuangan, khususnya pada transaksi kartu kredit, menyebabkan kerugian global mencapai miliaran dolar setiap tahunnya, dengan peningkatan kasus di Indonesia sebesar 18% pada tahun 2022 menurut laporan Bank Indonesia (2023). Tantangan utama dalam deteksi fraud terletak pada sifat data yang tidak seimbang (imbalanced), di mana kasus fraud hanya merupakan proporsi kecil dari total transaksi, sehingga model konvensional sering kali gagal mengidentifikasi pola mencurigakan secara akurat. Hal ini tidak hanya mengakibatkan kerugian finansial bagi individu dan lembaga keuangan, tetapi juga mengerosi kepercayaan publik terhadap sistem pembayaran digital.[2]

Untuk mengatasi isu tersebut, penerapan kecerdasan buatan, khususnya machine learning, menjadi solusi yang menjanjikan. Penelitian ini bertujuan mendeteksi fraud dengan teknik ensemble machine learning, khususnya Random Forest (RF) dan Extreme Gradient Boosting (XGBoost), diterapkan pada dataset Credit Card Fraud Detection Kaggle yang terdiri dari 284.807 sampel dengan hanya 0,17% kasus fraud. Pendekatan ensemble ini dipilih karena kemampuannya mengintegrasikan multiple algoritma untuk meningkatkan robustitas model. RF efektif mencegah overfitting lewat teknik bagging, sementara XGBoost unggul dalam optimasi gradient untuk data skala besar (Li & Chen, 2021). Proses metodologi mencakup preprocessing melalui normalisasi dan Synthetic Minority Over-sampling Technique (SMOTE) untuk menangani ketidakseimbangan, diikuti pelatihan model menggunakan Scikit-learn serta XGBoost di Python. Evaluasi dilakukan dengan metrik Area Under the Curve (AUC-ROC), precision, recall, dan F1-score melalui cross-validation, yang memungkinkan penilaian komprehensif terhadap performa model.

Hasil penelitian Al-Hashedi dan Magalingam (2021), menunjukkan bahwa pendekatan XGBoost lebih superior dibanding RF dalam hal akurasi, berkat ketangguhannya terhadap noise. Penelitian menyimpulkan XGBoost sebagai metode optimal untuk data imbalanced, dengan saran integrasi SMOTE guna tingkatkan recall hingga 15%. Kontribusi teoritis memperkaya literatur machine learning di keamanan siber (Al-Hashedi & Magalingam, 2021)[1], sementara manfaat praktisnya mengurangi kerugian bagi bank dan fintech, serta mendukung pengembangan model real-time. Dengan demikian, penelitian ini relevan untuk memperkuat keamanan siber di sektor keuangan Indonesia.

Sehingga berdasarkan permasalahan yang telah dijabarkan di atas, maka skripsi ini mengambi judul DETEKSI KECURANGAN KEUANGAN DENGAN MENGGUNAKAN METODE RANDOM FOREST DAN XGBOOST PADA DATA YANG TIDAKSEIMBANG. Diharapakan pendekatan yang digunakan dalam memperoleh hasil akurasi yang lebih baik dalam mendeteksi kecurang keuangan pada kartu kredit.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, penelitian ini merumuskan beberapa isu utama terkait deteksi kecurangan pada kartu kredit menggunakan machine learning. Rumusan masalah secara spesifik adalah sebagai berikut:

1. Bagaimana pengaruh penerapan teknik preprocessing, khususnya Synthetic Minority Over-sampling Technique (SMOTE), untuk meningkat kinerja akurasi pada model dalam mendeteksi kecurangan pada kartu kredit ?
2. Metode mana yang lebih optimal antara *Random Forest* dan XGBoost untuk deteksi kecurangan pada data transaksi keuangan, dengan mempertimbangkan aspek waktu komputasi dan ketahanan terhadap noise, dilihat dari metrik akurasi, precision, recall, F1-score, dan AUC-ROC ?

1.3 Batasan Masalah

Dalam penelitian ini, terdapat beberapa batasan yang perlu diperhatikan agar ruang lingkup dan fokus penelitian menjadi jelas serta hasil yang diperoleh

dapat dipertanggungjawabkan. Batasan-batasan tersebut adalah sebagai berikut:

1. Dataset yang digunakan adalah dataset historis statis dari Kaggle, yaitu dataset Credit Card Fraud Detection yang berisi data transaksi kartu kredit dengan label fraud dan non-fraud. Dataset ini bersifat anonim dan tidak mencakup data real-time atau data streaming yang dinamis, sehingga hasil penelitian ini tidak dapat langsung diaplikasikan pada sistem deteksi fraud secara real-time.
2. Metode yang digunakan terbatas pada dua algoritma ensemble machine learning, yaitu Random Forest (RF) dan Extreme Gradient Boosting (XGBoost). Penelitian ini tidak mengeksplorasi metode lain seperti deep learning, hybrid ensemble, atau teknik lain yang mungkin memiliki performa berbeda dalam mendeteksi fraud.
3. Hasil penelitian sangat bergantung pada kualitas dan karakteristik dataset Kaggle yang digunakan, yang mungkin tidak sepenuhnya mencerminkan variasi pola fraud di wilayah Indonesia atau skenario fraud terbaru setelah tahun 2023. Oleh karena itu, validasi lebih lanjut dengan data lokal atau data aktual dari lembaga keuangan sangat disarankan untuk penelitian lanjutan.
4. Preprocessing data yang dilakukan, seperti normalisasi dan penerapan Synthetic Minority Over-sampling Technique (SMOTE), difokuskan untuk mengatasi ketidakseimbangan data pada dataset ini. Namun, efektivitas teknik ini dapat berbeda jika diterapkan pada dataset dengan karakteristik yang berbeda.

1.4 Tujuan Penelitian

Adapun tujuan dilaksananakannya penelitian ini adalah sebagai berikut :

1. Untuk mengetahui bagaimana pengaruh penerapan teknik preprocessing, khususnya Synthetic Minority Over-sampling Technique (SMOTE), terhadap peningkatan performa model dalam mendeteksi kecurangan pada kartu kredit, terutama pada metrik recall.
2. Untuk mengetahui metode mana yang lebih optimal antara Random Forest dan XGBoost untuk deteksi fraud pada data transaksi keuangan, dengan

mempertimbangkan aspek waktu komputasi dan ketahanan terhadap noise, dilihat dari metrik AUC-ROC, precision, recall, dan F1-score melalui cross-validation.

1.6 Manfaat Penelitian

1. Manfaat Teknis

   a. Pengembangan Ilmu Pengetahuan: Memberi Perbandingan Kinerja Komprehensif Antar Algoritma penelitian ini menghasilkan evaluasi teknis tentang kinerja dua algoritma yang sangat populer untuk mendeteksi kecurangan, Random Forest dan XGBoost. Perbandingan ini membantu dalam menentukan algoritma mana yang lebih baik untuk digunakan dalam sistem deteksi kecurangan institusi keuangan.

   b. Panduan Teknis: Pengembangan model deteksi kecurangan keuangan yang mampu memberikan akurasi tinggi meskipun data yang digunakan tidak seimbang. Penggunaan algoritma Random Forest dan XGBoost terbukti dapat menangani pola kompleks pada data keuangan, yang membantu mengidentifikasi transaksi mencurigakan secara lebih efektif.

2. Manfaat Non Teknis

   a. Bagi Pengguna

      Mendapatkan keuntungan berupa perlindungan terhadap penipuan, sehingga transaksi secara digital menjadi lebih aman dan dapat dipercaya. Dengan sistem yang tepat, pengguna bisa menghindari kerugian uang pribadi karena penipuan, serta meningkatkan rasa percaya terhadap sistem pembayaran elektronik di Indonesia.

   b. Bagi Organisasi

      Membuat pengalaman pengguna lebih baik dan mengurangi rasa cemas karena ancaman keamanan siber.

c. Bagi Peneliti Selanjutnya

Menjadikan referensi dalam penelitian menggunakan dataset lokal Indonesia atau teknik hybrid lainnya, serta memperluas lingkup keamanan siber.

1.5 Sistematika Penulisan

Dalam penelitian ini, penulis membagi sistematika penulis menjadi beberapa bagian sesuai dengan permasalahan masi-masing sebagai berikut:

BAB I PENDAHULUAN, Membahas latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup, batasan, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA, Menyajikan landasan teori tentang kecurangan pada kartu kredit, machine learning, Random Forest, XGBoost, serta tinjauan penelitian terkait.

BAB III METODE PENELITIAN, Menguraikan desain penelitian, deskripsi dataset Kaggle, tahap preprocessing, implementasi model, dan prosedur evaluasi.

BAB IV HASIL DAN PEMBAHASAN, Menampilkan hasil eksperimen, perbandingan performa, serta analisis interpretatif terhadap temuan.

BAB V PENUTUP, Merangkum kesimpulan utama, implikasi praktis, serta rekomendasi untuk pengembangan lebih lanjut.