# **BAB III METODELOGI PENELITIAN**

## **Kerangka Kerja Penelitian**

Kerangka kerja penelitian ini dirancang untuk memberikan gambaran sistematis mengenai tahapan-tahapan dalam penerapan algoritma Random Forest guna mendeteksi potensi fraud pada setiap cabang perusahaan. Tahapan dimulai dari proses pengumpulan data, khususnya data yang mencerminkan indikator fraud berdasarkan Fraud Diamond Theory, dilanjutkan dengan pra-pemrosesan data seperti pembersihan, normalisasi, dan transformasi.

Setelah data siap, dilakukan pemilihan fitur (feature selection) untuk menentukan variabel-variabel yang paling berpengaruh dalam mendeteksi fraud. Selanjutnya, data dibagi menjadi data latih dan data uji, kemudian diolah menggunakan algoritma Random Forest untuk membangun model klasifikasi fraud dan non-fraud.

Model yang telah dibuat dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score guna memastikan keandalan prediksi. Hasil dari model ini digunakan untuk mengidentifikasi cabang-cabang perusahaan yang memiliki indikasi kuat melakukan tindakan fraud. Dengan kerangka kerja ini, diharapkan perusahaan dapat melakukan pencegahan dan pengawasan yang lebih efektif terhadap potensi kecurangan internal.

### **3.6.1 Desain Kerangka Kerja**

Desain kerangka kerja dalam penelitian ini disusun untuk menggambarkan alur sistematis dari proses pendeteksian fraud pada setiap cabang perusahaan menggunakan algoritma Random Forest. Rancangan ini mencakup seluruh tahapan penting yang dimulai dari proses pengumpulan data, pengolahan data, pelatihan model, evaluasi, hingga interpretasi hasil klasifikasi.

### **3.6.2 Alur Proses Deteksi Fraud**

Alur proses deteksi fraud dalam penelitian ini dirancang untuk menggambarkan secara sistematis setiap tahapan yang dilalui dalam mendeteksi indikasi kecurangan di masing-masing cabang perusahaan. Proses ini berbasis pada teori Fraud Diamond, yang menyatakan bahwa fraud terjadi karena adanya empat elemen utama: pressure (tekanan), opportunity (kesempatan), rationalization (pembenaran), dan capability (kemampuan). Keempat elemen ini digunakan sebagai dasar untuk menyusun indikator dalam pengumpulan data dan membangun model klasifikasi fraud menggunakan algoritma Random Forest. Adapun alur prosesnya dijelaskan sebagai berikut:

#### **3.6.2.1. Penyusunan Instrumen dan Pengumpulan Data**

Langkah awal dimulai dengan penyusunan kuisioner berdasarkan indikator-indikator dari keempat elemen dalam teori Fraud Diamond:

1. Pressure: misalnya kebutuhan ekonomi, target yang tinggi, atau utang pribadi.
2. Opportunity: lemahnya sistem pengawasan, celah dalam prosedur kerja.
3. Rationalization: pembenaran pribadi seperti "semua orang juga melakukannya".
4. Capability: kemampuan teknis dan otoritas yang dimiliki pelaku untuk melakukan fraud.

Kuisioner disebarkan kepada karyawan dari berbagai cabang perusahaan, lalu data dikumpulkan dalam bentuk numerik untuk dianalisis.

#### **3.6.2.2. Pra-pemrosesan Data**

Data mentah dari kuisioner selanjutnya diproses agar siap digunakan dalam model pembelajaran mesin. Tahapan ini mencakup:

1. Pembersihan data dari nilai kosong, duplikasi, dan kesalahan input.
2. Normalisasi atau standardisasi agar setiap fitur berada dalam skala yang seragam.
3. Pengkodean (jika ada variabel kategorikal).
4. Labeling: menetapkan apakah data termasuk kategori fraud atau non-fraud, berdasarkan skor total atau hasil validasi pakar.

#### **3.6.2.3. Seleksi Fitur (Feature Selection)**

Dari seluruh data yang dikumpulkan, dilakukan seleksi fitur untuk menentukan indikator mana yang paling relevan terhadap kejadian fraud. Seleksi ini penting untuk meningkatkan performa model dan mengurangi kompleksitas pemrosesan data.

#### **3.6.2.4. Pembagian Data**

Data dibagi menjadi dua bagian:

Data Latih (Training Set): digunakan untuk membangun model.

Data Uji (Testing Set): digunakan untuk menguji performa model.

Proporsi umum yang digunakan adalah 80:20. Alternatif lain, seperti k-fold cross-validation, dapat diterapkan untuk menghindari bias hasil evaluasi.

#### **3.6.2.5. Penerapan Algoritma Random Forest**

Model klasifikasi dibangun menggunakan algoritma Random Forest, yaitu metode ensemble yang menggabungkan banyak pohon keputusan (decision tree). Random Forest dipilih karena:

Mampu menangani banyak fitur.

Tahan terhadap overfitting.

Memberikan hasil klasifikasi yang akurat. Model dilatih menggunakan data latih dan parameter seperti jumlah pohon (n\_estimators) dan kedalaman pohon (max\_depth) disesuaikan melalui tuning.

### 3.6.3 Diagram Kerangka Kerja

Menampilkan diagram alur (flowchart atau blok diagram) untuk memvisualisasikan proses kerja dari awal hingga akhir.

## **Lokasi Riset (Jika ada) dan Sampel Data**

## **Waktu Pelaksanaan**